

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2004, 09:09:48 ; Search time 66 Seconds  
(without alignments)  
4096.939 Million cell updates/sec

Title: US-09-917-376-1  
Perfect score: 5135  
Sequence: 1 MDRSENIRLTMRSRRLVSLT.....RASFGSVNPATPTADTVLQX 957

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues  
Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: geneseqp1980s:\*

2: geneseqp1990s:\*

3: geneseqp2000s:\*

4: geneseqp2001s:\*

5: geneseqp2002s:\*

6: geneseqp2003as:\*

7: geneseqp2003bs:\*

8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5134	100.0	957	6	ABP73015
2	5134	100.0	957	7	ADD22821
3	4036	78.6	740	6	ABP73016
4	4036	78.6	740	7	ADD22926
5	4036	78.6	740	7	ADD22923
6	2174	42.3	940	6	ABP99489
7	1680	32.7	726	7	ADD22927
8	1625.5	31.7	838	7	ADD42055
9	848	16.5	555	7	ADD24919
10	823	16.0	1228	6	ABP73029
11	702	13.7	762	6	ABP73022
12	468	9.1	88	6	ABP73018
13	468	9.1	89	6	ABP73017
14	462	9.0	88	7	ADD22825
15	462	9.0	89	7	ADD22825
16	462	9.0	89	7	ADD22924
17	459	8.9	150	6	ABP73025
18	429	8.4	153	6	ABP71658
19	429	8.4	153	6	ABP71656
20	334	6.5	1751	5	AAV13493
21	334	6.5	1751	5	AAE16324
22	323	6.3	782	2	AAE15625
23	301.5	5.9	1749	4	AAE15625
24	288	5.6	96	5	ABP72314
25	276	5.4	1426	2	AAV13492

26	276	5.4	1426	5	AAE16323	AAE16323	Active ce
27	274	5.3	406	4	ABG27250	ABG27250	Novel hum
28	262	5.1	2468	6	ABU38411	ABU38411	Protein e
29	262	5.1	2468	6	ABP59933	ABP59933	Microbial
30	261.5	5.1	531	2	AAW15238	AAW15238	Scaffold1
31	261.5	5.1	1853	2	AAW43108	AAW43108	C. thermo
32	257.5	5.0	732	5	ABP93616	ABP93616	Herbicide
33	253	4.9	2435	6	ABU19529	ABU19529	Protein e
34	244	4.8	1664	2	AAW43106	AAW43106	C. thermo
35	243.5	4.7	1049	6	ABU21976	ABU21976	Protein e
36	243	4.7	1259	5	ABP57561	ABP57561	Human MC2
37	238	4.6	200	5	ABP78541	ABP78541	THP-Pro-T
38	237	4.6	103	5	ABP78540	ABP78540	Ala-Pro-A
39	234	4.6	2137	5	ABP39618	ABP39618	Staphyloc
40	232.5	4.5	1016	6	ABU41567	ABU41567	Protein e
41	229.5	4.5	950	4	ABP71513	ABP71513	Drosophila
42	227.5	4.4	455	4	AAU44346	AAU44346	Propionib
43	227.5	4.4	455	6	ABM40865	ABM40865	Protonib
44	227.5	4.4	2271	6	ABU16000	ABU16000	Protein e
45	226.5	4.4	1831	6	ABU43109	ABU43109	Protein e

ALIGNMENTS

RESULT 1  
ABP73015  
ID ABP73015 standard; protein; 957 AA.  
AC ABP73015;  
DT 03-JUN-2003 (first entry)  
XX  
XX  
XX Amino acid sequence of the avicelase AviiiI.  
XX Avicelase; AviiiI; glycoside hydrolase; enzyme; cellulase; biofuel;  
XX detergent; pulp processing; paper processing; feed processing; textile;  
XX cellulose.  
XX Acidothermus cellulolyticus.  
OS  
FH Key Location/Qualifiers  
FT Misc-difference 957  
FT /note="unspecified residue encoded by N"  
XX  
XX  
XX WO2003012090-A2.  
XX  
XX 13-FEB-2003.  
XX  
XX 28-JUL-2001; 2001WO-US023818.  
XX  
XX 28-JUL-2001; 2001WO-US023818.  
XX  
XX (MIDE ) MIDWEST RES INST.  
XX  
XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX WPI; 2003-248177/24.  
XX N-PDB; ABZ77632.  
XX  
XX New thermostable AviiiI peptide from Acidothermus cellulolyticus, useful  
XX for degradation of cellulose or in generating anti-AviiiI antibodies for  
XX purifying recombinant AviiiI polypeptides from genetically engineered  
XX host cells.  
XX  
XX Claim 2; Page 20; 4pp; English.  
XX  
XX The present sequence represents a thermostable avicelase polypeptide,  
XX designated AviiiI. AviiiI is a member of the glycoside hydrolase family  
XX of enzymes, and is a cellulase. AviiiI is useful in the conversion of  
XX biomass to biofuels and biofuel additives. It may be useful in the  
XX production of detergents, pulp and paper processing, food and feed  
XX processing and in textile processes. The thermostable AviiiI peptide is

CC useful in the degradation of cellulose, and in generating specific anti-  
 CC Aviii antibodies that are useful in purifying recombinant Aviii  
 CC polypeptides from genetically engineered host cells, in detecting Aviii  
 CC polypeptide expression, as well as a reagent tool for characterizing the  
 CC molecular actions of the polypeptide. The Aviii polynucleotide is useful  
 CC as a source of probes or primers in various diagnostic assays

XX  
 SQ Sequence 957 AA;

Query Match 100.0%; Score 5134; DB 6; Length 957;  
 Best Local Similarity 100.0%; Pred. No. 4e-263;  
 Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDRSENRLTMRSRRLVSLAATASFAVAALGVPLAIFASPAHAATTOPYTWSNVAIG 60  
 DB 1 MDRSENRLTMRSRRLVSLAATASFAVAALGVPLAIFASPAHAATTOPYTWSNVAIG 60  
 QY 61 GGGFVDGIVFNEGAPGILYVRTDIGMYRMDANGRMIPLDWVGNNWGVVSIAD 120  
 DB 61 GGGFVDGIVFNEGAPGILYVRTDIGMYRMDANGRMIPLDWVGNNWGVVSIAD 120  
 QY 121 PINNKVMAAVGNTNSWDNDGAILRSSDQATWQITPLPKLGNNMGRGMBRLAD 180  
 DB 121 PINNKVMAAVGNTNSWDNDGAILRSSDQATWQITPLPKLGNNMGRGMBRLAD 180  
 QY 181 PNNNTILYFGAPSGKGLWRSTDSGATWSOMTFPDVGYIANPTDTTQSDIQGVVWA 240  
 DB 181 PNNNTILYFGAPSGKGLWRSTDSGATWSOMTFPDVGYIANPTDTTQSDIQGVVWA 240  
 QY 241 FDKSSSLIGQASKTIFVGADPNNFVFWSRDGAATWQAVGAPGTGFIPIKGVDFPNVHL 300  
 DB 241 FDKSSSLIGQASKTIFVGADPNNFVFWSRDGAATWQAVGAPGTGFIPIKGVDFPNVHL 300  
 QY 241 FDKSSSLIGQASKTIFVGADPNNFVFWSRDGAATWQAVGAPGTGFIPIKGVDFPNVHL 300  
 DB 241 FDKSSSLIGQASKTIFVGADPNNFVFWSRDGAATWQAVGAPGTGFIPIKGVDFPNVHL 300  
 QY 301 YIATSNNGGPYDSSGDVWKFSTVSTGTRISPVSTPDANDYFCGSGLTIRQHNNTIM 360  
 DB 301 YIATSNNGGPYDSSGDVWKFSTVSTGTRISPVSTPDANDYFCGSGLTIRQHNNTIM 360  
 QY 361 VAFQISWMPDITIRSTDGATWTRIMWTSYPNSLRVLDISAEPMILFVGQPPVP 420  
 DB 361 VAFQISWMPDITIRSTDGATWTRIMWTSYPNSLRVLDISAEPMILFVGQPPVP 420  
 QY 421 SPKLGWMDENALIDPFNSDRMLYGTGATLYATNDLTKMDSGQIHIAIPVKGLEBTAVND 480  
 DB 421 SPKLGWMDENALIDPFNSDRMLYGTGATLYATNDLTKMDSGQIHIAIPVKGLEBTAVND 480  
 QY 481 LISPPSGAPLISALGDLGFTHADYTAVPSTIFTSVPFTTGSVDYAEINPSIIVAGSF 540  
 DB 481 LISPPSGAPLISALGDLGFTHADYTAVPSTIFTSVPFTTGSVDYAEINPSIIVAGSF 540  
 QY 541 DPSSQPNDRHVAFTSDGKMMFQSGEPGVTTGTVAAADSRFVWAGDPQPPVAV 600  
 DB 541 DPSSQPNDRHVAFTSDGKMMFQSGEPGVTTGTVAAADSRFVWAGDPQPPVAV 600  
 QY 601 GFSGNSWASQGVPAQAQRSDRVNPKTFYALNGTFYRSTDGVTQFPVAAAGLPSSGAVG 660  
 DB 601 GFSGNSWASQGVPAQAQRSDRVNPKTFYALNGTFYRSTDGVTQFPVAAAGLPSSGAVG 660  
 QY 661 VFMAVPGKEGDLWLAASGLYHSTNGSSWSAITGVSSAVVGFKSAFGSSYPVAVV 720  
 DB 661 VFMAVPGKEGDLWLAASGLYHSTNGSSWSAITGVSSAVVGFKSAFGSSYPVAVV 720  
 QY 721 GTIGVTVGAYRSDDCGTTWVLINDOHOYGMNGQALTGHNALRRYITGNGGIYYGDI 780  
 DB 721 GTIGVTVGAYRSDDCGTTWVLINDOHOYGMNGQALTGHNALRRYITGNGGIYYGDI 780  
 QY 781 GGASGSGSP 840  
 DB 781 GGASGSGSP 840  
 QY 841 SASPSPPSSSPSPSSSP 900  
 DB 841 SASPSPPSSSPSPSSSP 900

QY 901 SSVDLSTVYRKYFTRDGSSTLYVNCDMAAIGCNRASFGSVNPTPTADTYLQ 956  
 DB 901 SSVDLSTVYRKYFTRDGSSTLYVNCDMAAIGCNRASFGSVNPTPTADTYLQ 956

# RESULT 2

ADD22921  
 ID ADD22921 standard; protein; 957 AA.

AC ADD22921;

DT 15-JUN-2004 (first entry)

XX Acidothermus cellulolyticus avicelase Aviii.

XX enzyme; Aviii; cellulose reduction; agricultural biomass;

XX municipal solid waste; glycoside hydrolase; avicelase.

XX Acidothermus cellulolyticus.

FT Key Location/Qualifiers

FT Misc-difference 957 /label= Unknown

XX /note= "Encoded by N"

PD US2003108988-A1.

XX 12-JUN-2003.

XX 18-OCT-2002; 2002US-00155400.

XX 28-JUL-2001; 2001US-00917376.

XX (DING/) DING S.

PA (ADNE/) ADNEY W S.

PA (VINZ/) VINZANT T B.

PA (HIMM/) HIMMEL M E.

PI Ding S, Adney WS, Vinzant TB, Himmel ME;

XX WPI; 2003-810853/76.

DR N-PSDB; ADD22922.

XX New isolated thermal tolerant avicelase polynucleotide useful for

PT detection of a polynucleotide encoding Aviii and for reducing cellulose

PT in a starting material, e.g. municipal solid waste.

XX Claim 16; SEQ ID NO 1; 29pp; English.

XX The invention relates to an isolated polynucleotide molecule encoding a

CC thermostable Aviii polypeptide. The polynucleotide is useful for

CC detection of a polynucleotide encoding Aviii. The polynucleotide is

CC useful for reducing cellulose in a starting material which involves

CC administering to the starting material, e.g. agricultural biomass or

CC municipal solid waste, a polypeptide molecule of the polynucleotide. The

CC method further comprises administering a second polypeptide molecule

CC chosen from the glycoside hydrolase family of proteins. The present

CC sequence represents the amino acid sequence of Acidothermus

CC cellulolyticus avicelase Aviii.

SQ Sequence 957 AA;

Query Match 100.0%; Score 5134; DB 7; Length 957;

Best Local Similarity 100.0%; Pred. No. 4e-263;

Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDRSENRLTMRSRRLVSLAATASFAVAALGVPLAIFASPAHAATTOPYTWSNVAIG 60

DB 1 MDRSENRLTMRSRRLVSLAATASFAVAALGVPLAIFASPAHAATTOPYTWSNVAIG 60

QY 61 GGGFVDGIVFNEGAPGILYVRTDIGMYRMDANGRMIPLDWVGNNWGVVSIAD 120

DB 61 GGGFVDGIVFNEGAPGILYVRTDIGMYRMDANGRMIPLDWVGNNWGVVSIAD 120

QY 121 PINTNKVAAVGMVYNSWDPNDAILRSSDQATWQITPLPKLGANNPGRMGERLAVD 180  
DB 121 PINTNKVAAVGMVYNSWDPNDAILRSSDQATWQITPLPKLGANNPGRMGERLAVD 180  
QY 181 PNNNNILYFGAPSGKGLMRSTDSGATWQMTNFPVGVYIANPPTTGGQSDIQGVVWVA 240  
DB 181 PNNNNILYFGAPSGKGLMRSTDSGATWQMTNFPVGVYIANPPTTGGQSDIQGVVWVA 240  
QY 241 FDKSSSSIGQASKTIFVGVADPNNPFWSRDQATWQAVPAGPTGFI PHKGVFDPVNVHL 300  
DB 241 FDKSSSSIGQASKTIFVGVADPNNPFWSRDQATWQAVPAGPTGFI PHKGVFDPVNVHL 300  
QY 301 YIATSNTPGPDYSGSDVWKFVSITGVTIRISPVSTDTANDYFGYSGLTIDRQHPNTIM 360  
DB 301 YIATSNTPGPDYSGSDVWKFVSITGVTIRISPVSTDTANDYFGYSGLTIDRQHPNTIM 360  
QY 361 VATQISWMPDITIFRSTDTGATWTRIMWTSYPNRSILRYLDISAPWLTFTGVQNPVPV 420  
DB 361 VATQISWMPDITIFRSTDTGATWTRIMWTSYPNRSILRYLDISAPWLTFTGVQNPVPV 420  
QY 421 SPKLGMDDEAALIDPNSDRMLYGTATLYATNDLTKWDSGQIHIAPVVKLEETAVND 480  
DB 421 SPKLGMDDEAALIDPNSDRMLYGTATLYATNDLTKWDSGQIHIAPVVKLEETAVND 480  
QY 481 LISPPSGAPLISALGDLGFTHADVTAVPSTIFTSPTVTTGTSVDYAEINPISIIYRAGSF 540  
DB 481 LISPPSGAPLISALGDLGFTHADVTAVPSTIFTSPTVTTGTSVDYAEINPISIIYRAGSF 540  
QY 541 DPSSQPNDRHVAFTDGGKNNFQSGEPGVTTGGTVAASADGSRFVMAFGDPGQPVVAV 600  
DB 541 DPSSQPNDRHVAFTDGGKNNFQSGEPGVTTGGTVAASADGSRFVMAFGDPGQPVVAV 600  
QY 601 GFGNSMAASQGVPAQAQIRSDRVNPKTFYALSNGFYRSTGGVTFQPVAAQLPSSGAVG 660  
DB 601 GFGNSMAASQGVPAQAQIRSDRVNPKTFYALSNGFYRSTGGVTFQPVAAQLPSSGAVG 660  
QY 661 VMFAVAVGKEGDLMLAASSGLYHSTNGSSWSAITGVSSAAVAVGKAPGSSYPAVAV 720  
DB 661 VMFAVAVGKEGDLMLAASSGLYHSTNGSSWSAITGVSSAAVAVGKAPGSSYPAVAV 720  
QY 721 GTTGGVTVGAYRSDCGTTWVLINDQHQYGNMGQAITGDHANLRRVYI GTNGRGIVYGD 780  
DB 721 GTTGGVTVGAYRSDCGTTWVLINDQHQYGNMGQAITGDHANLRRVYI GTNGRGIVYGD 780  
QY 781 GGAPSGSPSPSVSASPSLSPPSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSP 840  
DB 781 GGAPSGSPSPSVSASPSLSPPSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSP 840  
QY 841 SASPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSS 900  
DB 841 SASPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSS 900  
QY 901 SSVDLSVTVRYWTRTGGSSSTLYVNCMAAIGCNIRASFGSVNPAITPDTYIQ 956  
DB 901 SSVDLSVTVRYWTRTGGSSSTLYVNCMAAIGCNIRASFGSVNPAITPDTYIQ 956

RESULT 3  
ABP73016  
ID ABP73016 standard; peptide; 740 AA.

XX 03-JUN-2003 (first entry)  
XX AC ABP73016;  
XX DE Amino acid sequence of the avicelase AvilIII catalytic domain.  
XX KW Avicelase; AvilIII; glycoside hydrolase; enzyme; cellulase; biofuel;  
XX KM detergent; pulp processing; paper processing; feed processing; textile;  
XX OS cellulose.  
XX Acidothermus cellulyticus.

XX XH  
PN WO2003012090-A2.  
XX PD 13-FEB-2003.  
XX PF 28-JUL-2001; 2001MO-US023818.  
XX PR 28-JUL-2001; 2001MO-US023818.  
XX PA (MIDE ) MIDWEST RES INST.  
XX PI Ding S, Adney WS, Vinzant TB, Himmel ME.  
XX DR WPI; 2003-248177/24.  
XX PT New thermostable AvilIII peptide from Acidothermus cellulyticus, useful  
PT for degradation of cellulose or in generating anti-AvilIII antibodies for  
PT purifying recombinant AvilIII polypeptides from genetically engineered  
PT host cells.  
XX PS Claim 6; Page 8; 44pp; English.  
XX CC The present sequence is derived from a thermostable avicelase, designated  
CC AvilIII. AvilIII is a member of the glycoside hydrolase family of enzymes,  
CC and is a cellulase. AvilIII is useful in the conversion of biomass to  
CC biofuels and biofuel additives. It may be useful in the production of  
CC detergents, pulp and paper processing, food and feed processing and in  
CC textile processes. The thermostable AvilIII peptide is useful in the  
CC degradation of cellulose, and in generating specific anti-AvilIII  
CC antibodies that are useful in purifying recombinant AvilIII polypeptides  
CC from genetically engineered host cells, in detecting AvilIII polypeptide  
CC expression, as well as a reagent tool for characterizing the molecular  
CC actions of the polypeptide. The AvilIII polynucleotide is useful as a  
CC source of probes or primers in various diagnostic assays  
XX SQ Sequence 740 AA;  
Query Match 78.6%; Score 4036; DB 6; Length 740;  
Best Local Similarity 100.0%; Pred. No. 3.2e-205;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 47 ATTQPTWNSVNAIGGGFVDGIVFNEGAGILYVRDIDGMYMDANGRMWILBLDWGM 106  
DB 1 ATTQPTWNSVNAIGGGFVDGIVFNEGAGILYVRDIDGMYMDANGRMWILBLDWGM 60  
QY 107 NNWGVNVSIADPINTNKVAAVGMVYNSWDPNDAILRSSDQATWQITPLPKLG 166  
DB 61 NNWGVNVSIADPINTNKVAAVGMVYNSWDPNDAILRSSDQATWQITPLPKLG 120  
QY 167 NMPGRMGERLAVDPNNNDILYFGAPSGKGLMRSTDSGATWQMTNFPVGVYIANPDT 226  
DB 121 NMPGRMGERLAVDPNNNDILYFGAPSGKGLMRSTDSGATWQMTNFPVGVYIANPDT 180  
QY 227 TGYQSDIQGVVWVAFPKSSSSIGQASKTIFVGVADPNNPFWSRDQATWQAVPAGPTGF 286  
DB 181 TGYQSDIQGVVWVAFPKSSSSIGQASKTIFVGVADPNNPFWSRDQATWQAVPAGPTGF 240  
QY 287 IPHKGVFDPVNVHLYIATSNTPGPDYSGSDVWKFVSITGVTIRISPVSTDTANDYFGY 346  
DB 241 IPHKGVFDPVNVHLYIATSNTPGPDYSGSDVWKFVSITGVTIRISPVSTDTANDYFGY 300  
QY 347 SGLTIDRQHPNTIMVATQISWMPDITIFRSTDTGATWTRIMWTSYPNRSILRYLDISAB 406  
DB 301 SGLTIDRQHPNTIMVATQISWMPDITIFRSTDTGATWTRIMWTSYPNRSILRYLDISAB 360  
QY 407 PMLTFTGVQNPVPVPSKLGMDDEAALIDPNSDRMLYGTATLYATNDLTKMDSGQIH 466  
DB 361 PMLTFTGVQNPVPVPSKLGMDDEAALIDPNSDRMLYGTATLYATNDLTKMDSGQIH 420  
QY 467 APMVVKLEETAVNDLISPPSGAPLISALGDLGFTHADVTAVPSTIFTSPTVTTGTSVDY 526  
DB 421 APMVVKLEETAVNDLISPPSGAPLISALGDLGFTHADVTAVPSTIFTSPTVTTGTSVDY 480

QY 527 AELNPSIIIVRAGSDPSSQPNDRHVAFTDGGKWMFGSGEGVTTGGTVAASADGSRFV 586  
| | | | |  
DB 481 AELNPSIIIVRAGSDPSSQPNDRHVAFTDGGKWMFGSGEGVTTGGTVAASADGSRFV 540  
| | | | |  
QY 587 WAPGDPGPVVVYAVGFGNSMAASQGVPPANAQIRSDRVNPKTFYALNSTEYFRSTDCGVT 646  
| | | | |  
DB 541 WAPGDPGPVVVYAVGFGNSMAASQGVPPANAQIRSDRVNPKTFYALNSTEYFRSTDCGVT 600  
| | | | |  
QY 647 QPVAAGLPSSGAVGMFHAVPKSGDLMLAASGLYHSTNGGSSWSAITGVSSAVNVFG 706  
| | | | |  
DB 601 QPVAAGLPSSGAVGMFHAVPKSGDLMLAASGLYHSTNGGSSWSAITGVSSAVNVFG 660  
| | | | |  
QY 707 KSAFGSSYPVAVFVGTIGGVTGAYRSDDCGTTWVLLINDQHQYGNWGOAITGDHANLRV 766  
| | | | |  
DB 661 KSAFGSSYPVAVFVGTIGGVTGAYRSDDCGTTWVLLINDQHQYGNWGOAITGDHANLRV 720  
| | | | |  
QY 767 YIGTNGRGIVYGDIGGAPSG 786  
| | | | |  
DB 721 YIGTNGRGIVYGDIGGAPSG 740  
| | | | |

## RESULT 4

ADD22926 ID ADD22926 standard; protein; 740 AA.

ADD22926;

15-JAN-2004 (first entry)

Acidothermus cellulolyticus avicelase AvIII catalytic domain.

enzyme; AvIII; cellulose reduction; agricultural biomass;

municipal solid waste; glycoside hydrolase; avicelase.

Acidothermus cellulolyticus.

US2003108988-A1.

12-JUN-2003.

18-OCT-2002; 2002US-00155400.

28-JUL-2001; 2001US-00917376.

(DING/) DING S.

(ADNE/) ADNEY W S.

(VINZ/) VINZANT T B.

(HIMM/) HIMMEL M E.

Ding S, Adney WS, Vinzant TB, Himmel ME;

WPI; 2003-810853/76.

New isolated thermal tolerant avicelase polynucleotide useful for  
detection of a polynucleotide encoding AvIII and for reducing cellulose  
in a starting material, e.g. municipal solid waste.

Example 2; SEQ ID NO 6; 29pp; English.

The invention relates to an isolated polynucleotide molecule encoding a  
thermostable AvIII polypeptide. The polynucleotide is useful for  
detection of a polynucleotide encoding AvIII. The polynucleotide is  
useful for reducing cellulose in a starting material which involves  
administering to the starting material, e.g. agricultural biomass or  
municipal solid waste, a polypeptide molecule of the polynucleotide. The  
method further comprises administering a second polypeptide molecule  
chosen from the glycoside hydrolase family of proteins. The present  
sequence represents the amino acid sequence of Acidothermus  
cellulolyticus avicelase AvIII catalytic domain.

Sequence 740 AA;

Query Match 78.6%; Score 4036; DB 7; Length 740;

Best Local Similarity 100.0%; Pred. No. 3.2e-205; Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 ATTQPYTMSNVAIGGGFVDGIVFNEGAPGIIYVRTIDIGMYRWDAANGRMIPLDWMG 106  
| | | | |  
DB 1 ATTQPYTMSNVAIGGGFVDGIVFNEGAPGIIYVRTIDIGMYRWDAANGRMIPLDWMG 60  
| | | | |  
QY 107 NNMGVGVSIADPINTNKVAAVGMVNTSMDPNDGAILRSSDQATWQITPLPKLG 166  
| | | | |  
DB 61 NNMGVGVSIADPINTNKVAAVGMVNTSMDPNDGAILRSSDQATWQITPLPKLG 120  
| | | | |  
QY 167 NNPGKMGERRLAVDPPNDNIIYFGAPSSGKLMRSTDSGATWSQMTNPPVGYTIANPT 226  
| | | | |  
DB 121 NNPGKMGERRLAVDPPNDNIIYFGAPSSGKLMRSTDSGATWSQMTNPPVGYTIANPT 180  
| | | | |  
QY 227 TGYOSDIOGVWVAFPKSSSSIGQASKTIFVGADPNNVFMSRDGATWQAVGAPTG 286  
| | | | |  
DB 181 TGYOSDIOGVWVAFPKSSSSIGQASKTIFVGADPNNVFMSRDGATWQAVGAPTG 240  
| | | | |  
QY 287 IPHKGVPDPVNHVLYATSNTPGYPYDSSGQDWKFSVTSQWTRISPVSTDPANDYFG 346  
| | | | |  
DB 241 IPHKGVPDPVNHVLYATSNTPGYPYDSSGQDWKFSVTSQWTRISPVSTDPANDYFG 300  
| | | | |  
QY 347 SGLTIDRQHPNTIMVATQISWEPDIIIPRSTGATWTRIMWTSYIPNRLRYVDISAE 406  
| | | | |  
DB 301 SGLTIDRQHPNTIMVATQISWEPDIIIPRSTGATWTRIMWTSYIPNRLRYVDISAE 360  
| | | | |  
QY 407 PMLTFGVQNPVPVPSKLGMDDEAMAIDPNSDRMLYGTATLYATNDLTKMDSGQIHT 466  
| | | | |  
DB 361 PMLTFGVQNPVPVPSKLGMDDEAMAIDPNSDRMLYGTATLYATNDLTKMDSGQIHT 420  
| | | | |  
QY 467 APWVGLGETAVNDLISPPSAPLISALGDLGFTHADTAVAPSTIFTSVFETGTSVDY 526  
| | | | |  
DB 421 APWVGLGETAVNDLISPPSAPLISALGDLGFTHADTAVAPSTIFTSVFETGTSVDY 480  
| | | | |  
QY 527 AELNPSIIIVRAGSDPSSQPNDRHVAFTDGGKWMFGSGEGVTTGGTVAASADGSRFV 586  
| | | | |  
DB 481 AELNPSIIIVRAGSDPSSQPNDRHVAFTDGGKWMFGSGEGVTTGGTVAASADGSRFV 540  
| | | | |  
QY 587 WAPGDPGPVVVYAVGFGNSMAASQGVPPANAQIRSDRVNPKTFYALNSTEYFRSTDCGVT 646  
| | | | |  
DB 541 WAPGDPGPVVVYAVGFGNSMAASQGVPPANAQIRSDRVNPKTFYALNSTEYFRSTDCGVT 600  
| | | | |  
QY 647 QPVAAGLPSSGAVGMFHAVPKSGDLMLAASGLYHSTNGGSSWSAITGVSSAVNVFG 706  
| | | | |  
DB 601 QPVAAGLPSSGAVGMFHAVPKSGDLMLAASGLYHSTNGGSSWSAITGVSSAVNVFG 660  
| | | | |  
QY 707 KSAFGSSYPVAVFVGTIGGVTGAYRSDDCGTTWVLLINDQHQYGNWGOAITGDHANLRV 766  
| | | | |  
DB 661 KSAFGSSYPVAVFVGTIGGVTGAYRSDDCGTTWVLLINDQHQYGNWGOAITGDHANLRV 720  
| | | | |  
QY 767 YIGTNGRGIVYGDIGGAPSG 786  
| | | | |  
DB 721 YIGTNGRGIVYGDIGGAPSG 740  
| | | | |

## RESULT 5

ADD22923 ID ADD22923 standard; protein; 740 AA.

ADD22923;

15-JAN-2004 (first entry)

Acidothermus cellulolyticus avicelase AvIII catalytic domain.

enzyme; AvIII; cellulose reduction; agricultural biomass;

municipal solid waste; glycoside hydrolase; avicelase.

Acidothermus cellulolyticus.

US2003108988-A1.



PD 12-JUN-2003.  
 XX PF 18-OCT-2002; 2002US-00155400.  
 XX PR 28-JUL-2001; 2001US-00917376.  
 XX (DING/) DING S.  
 PA (ADNE/) ADNEY W S.  
 PA (VIN/) VINZANT T B.  
 PA (HIMM/) HIMMEL M E.  
 PI Ding S, Adney WS, Vinzant TB, Himmel ME;  
 DX WPI; 2003-810853/76.  
 XX  
 PT New isolated thermal tolerant avicelase polynucleotide useful for  
 PT detection of a polynucleotide encoding Aviii and for reducing cellulose  
 PT in a starting material, e.g. municipal solid waste.  
 XX  
 XX Claim 16; SEQ ID NO 3; 29pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide molecule encoding a  
 CC thermostable Aviii polypeptide. The polynucleotide is useful for  
 CC detection of a polynucleotide encoding Aviii. The polynucleotide is  
 CC useful for reducing cellulose in a starting material which involves  
 CC administering to the starting material, e.g. agricultural biomass or  
 CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
 CC method further comprises administering a second polypeptide molecule  
 CC chosen from the glycoside hydrolase family of proteins. The present  
 CC sequence represents the amino acid sequence of Acidothermus  
 CC cellulolyticus avicelase Aviii catalytic domain.  
 XX  
 XX Sequence 740 AA;  
 SO  
 Query Match 78.6%; Score 4036; DB 7; Length 740;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-205;  
 Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 47 ATTGPTWMSNVAIGGGGFDGIVNEGAPGLIYRTDGGYRMDAANGRWIPLDWMGM 106  
 DB 1 ATTGPTWMSNVAIGGGGFDGIVNEGAPGLIYRTDGGYRMDAANGRWIPLDWMGM 60  
 QY 107 NNMGYNGVSIADPINTNKMAAVGMVYTNMWDNDGAILRSSDGAATWQITPLPKLGG 166  
 DB 61 NNMGYNGVSIADPINTNKMAAVGMVYTNMWDNDGAILRSSDGAATWQITPLPKLGG 120  
 QY 167 NMEGRGMEKRLAVDPNNNDNIIYFGAPSGKILMRSTDGATWSQMTNPFDPVGTIYIANPDT 226  
 DB 121 NMEGRGMEKRLAVDPNNNDNIIYFGAPSGKILMRSTDGATWSQMTNPFDPVGTIYIANPDT 180  
 QY 227 TGYOSDIQGVVWVAFDKSSSLGQASKTIFGVADPNNPVMSRPGCATWQAVPAPGTF 286  
 DB 181 TGYOSDIQGVVWVAFDKSSSLGQASKTIFGVADPNNPVMSRPGCATWQAVPAPGTF 240  
 QY 287 IPRKGVDPNNVHLYIATSNWTCGYDSSGDMWKFVSVTGWTIRISPPSTDTANDYEGY 346  
 DB 241 IPRKGVDPNNVHLYIATSNWTCGYDSSGDMWKFVSVTGWTIRISPPSTDTANDYEGY 300  
 QY 347 SGLTIDRQHNNTIMVATQISWMPDTIIFRSTDGATWTRIMDWTSYPRNSLRVYLDISAE 406  
 DB 301 SGLTIDRQHNNTIMVATQISWMPDTIIFRSTDGATWTRIMDWTSYPRNSLRVYLDISAE 360  
 QY 407 PMLTFCGQPNRPVPSFKLGMDMDMAIDPFNSDKMLYGTGATLVATNDLTKWDSGQIHI 466  
 DB 361 PMLTFCGQPNRPVPSFKLGMDMDMAIDPFNSDKMLYGTGATLVATNDLTKWDSGQIHI 420  
 QY 467 APWVKGLAEYAVNDLISPPSGAPLISALGDTGFTHADTVTPSTIFTSPTTGTSTVY 526  
 DB 421 APWVKGLAEYAVNDLISPPSGAPLISALGDTGFTHADTVTPSTIFTSPTTGTSTVY 480  
 QY 527 AELNPSIIVRAGSPSSQPNDRHVAFTSDGKWMFGQSEPGVTTGGTVAASADGSRFV 586  
 DB 481 AELNPSIIVRAGSPSSQPNDRHVAFTSDGKWMFGQSEPGVTTGGTVAASADGSRFV 540

QY 587 WAPDFGQPVYTVAVFGGNSMAASQGVPAANAQIRSDRVNPKTFYALSNGTFRSTDGCTF 646  
 DB 541 WAPDFGQPVYTVAVFGGNSMAASQGVPAANAQIRSDRVNPKTFYALSNGTFRSTDGCTF 600  
 QY 647 QPVAAGLPSSGAVGVVHFHVPKESGDLMLAASGLHSTWGGSSMSAITGVSAVAVGFG 706  
 DB 601 QPVAAGLPSSGAVGVVHFHVPKESGDLMLAASGLHSTWGGSSMSAITGVSAVAVGFG 660  
 QY 707 KSAFGSSYPVAVFVGTIGGVTGAYRSDDCGTTWVLINDOHOYGMCOAITGDHANIRRV 766  
 DB 661 KSAFGSSYPVAVFVGTIGGVTGAYRSDDCGTTWVLINDOHOYGMCOAITGDHANIRRV 720  
 QY 767 YIGTNGRGIVYGDIGAPSG 786  
 DB 721 YIGTNGRGIVYGDIGAPSG 740  
 RESULT 6  
 ABB99489  
 ID ABB99489 standard; protein; 940 AA.  
 XX  
 AC ABB99489;  
 XX  
 DT 03-MAR-2003 (first entry)  
 XX  
 DE Amino acid sequence of a xyloglucanase enzyme.  
 XX  
 KM Xyloglucanase; enzyme; family 74; glycosyl hydrolase; cellulose fiber;  
 KM textile scouring; cellulose fiber processing; ratting.  
 XX  
 OS Jonesia sp.  
 XX  
 PN WO200277242-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-DK000210.  
 XX  
 PR 27-MAR-2001; 2001DK-00000504.  
 XX  
 PA (NOVO ) NOVOZIMES AS.  
 XX  
 PI Duffner F, Sjoelund C;  
 XX  
 DR WPI: 2003-092855/08.  
 XX  
 DR N-PSDB; ABV76941.  
 XX  
 PT New xyloglucanase enzyme belonging to family 74 of glycosyl hydrolases,  
 PT and endogenous to a bacterium, useful in the textile industry for  
 PT improving properties of cellulosic fibers, yarn or fabric.  
 XX  
 PS Claim 5; Page 72-75; 76pp; English.  
 XX  
 CC The present sequence represents a xyloglucanase enzyme, belonging to  
 CC family 74 of glycosyl hydrolases. The enzyme is isolated from Jonesia sp.  
 CC DSM4140. The enzyme is useful in processes for machine treatment of  
 CC fabrics. It is also useful in the textile industry for improving the  
 CC properties of cellulosic fibers, yarn, woven or non-woven fabric, and in  
 CC a textile scouring process step. The xyloglucanase enzyme is also useful  
 CC in the cellulose fiber processing industry for ratting of fibers such as  
 CC hemp, jute, flax and linen. It is useful for preventing binding of  
 CC certain soils to the xyloglucan left on the cellulosic material  
 XX  
 SO Sequence 940 AA;  
 Query Match 42.3%; Score 2174; DB 6; Length 940;  
 Best Local Similarity 44.8%; Pred. No. 1.1e-106;  
 Matches 430; Conservative 141; Mismatches 311; Indels 78; Gaps 16;  
 QY 10 TMSRRLVSLMATAFAVAALGVLPITATSPAHAAATQPTWMSNVAIGGGFVDGIV 69  
 DB 5 TVVATLTLTLALTTAGIAATG---TPMEVTAPSPSSPSGYSWNVETVGGFVPGIV 61



Db 409 KANVTASLAVGIEEMAVLGLITPPGPAALLSAVGDDGGFTHSDDAPMNAIYHTPTTGT 468  
QY 521 GTSVDAELNPSIIYRAGSFDPSSQPNDRHVAFTDGGKMFQSGEPGVTGTGTVAA5A 580  
Db 469 TNGIDYAGNKSNTIVRSGASD--DYPT---LALSSNPGSTWADVAATSTGTGVALISA 523  
QY 581 DGSFVWAPGDPGPVYVAVFGNSMAASQGVANAAQIRSDRVNPKTYALSNGTFFNST 640  
Db 524 DGDITVLMSSSTGALVSKSGQ---TLTAVSLIPSGAVIASDKSDMTVFEGSGAGAIYSK 580  
QY 641 DGGTTPQVPAAGLSSGAVGVMFHVPKKEGDLMLAASSGLYHSTNGSSWSAI--TGVSS 699  
Db 581 NTATSTFKTVS-LGSSSTTVNAI-RAHPSIAGDVMASTDKGLMHSSTDYSGSTFTQISGVYA 638  
QY 700 AVNVGFKSAPGSSYPFAVFGVGTIGVAGAYRSDCGTTLVINDOHOYGNWGAQ--IT 757  
Db 639 GWSGTFKASSTGTVYVYGFFTIDGAAGLFRSEIAGNTQVYISDASHGFGS--GSANVYN 697  
QY 758 GDHANLRVYIIGTNGRGIVYDIGAPSG 786  
Db 698 GDLOTYGRVFRGHRPGLRLRQSGRFPAG 726  
RESULT 8  
ID ADD42055 standard; protein; 838 AA.  
XX ADD42055;  
XX  
XX 15-JAN-2004 (first entry)  
DE Trichoderma reesei Family 74 xyloglucanase.  
XX  
KW Trichoderma reesei; Family 74 xyloglucanase; beta-1,4-glycosidic linkage;  
KM xyloglucan hydrolases; biomass degradation; cellulose; hemicellulose;  
KW ethanol production; detergent composition; fabric treatment;  
KM textile treatment; enzyme.  
XX  
OS Hypocrea jecorina.  
XX  
PN WO2003089598-A2.  
PD 30-OCT-2003.  
XX  
PF 17-APR-2003; 2003WO-US011831.  
XX  
PR 19-APR-2002; 2002US-0373987P.  
XX  
PA (NOVO ) NOVOZYMES BIOTECH INC.  
XX  
PI Michael R, Zaretsky E, Haas J;  
XX  
XX WPI; 2003-845528/78.  
DR N-PSDB; ADD42054, ADD42060.  
XX  
PT New polypeptides having Family 74 xyloglucanase activity, and encoding  
XX nucleic acid molecules, useful for degrading cellulose- and hemicellulose  
XX -containing biomass to ethanol or as a detergent.  
PS Claim 8; SEQ ID NO 2; 96pp; English.  
XX  
CC The invention relates to a Family 74 xyloglucanase (ADD42055) from the  
CC fungus Trichoderma reesei (Hypocrea jecorina), and nucleic acids encoding  
CC it (e.g., ADD42054). The enzyme hydrolyses the beta-1,4-glycosidic  
CC linkages in the backbone of xyloglucan to xyloglucan oligosaccharides.  
CC The invention also relates polypeptide sequences at least 70% identical  
CC to the enzyme, expression vectors and host cells comprising a nucleic  
CC acid of the invention, the recombinant production of the enzyme, and  
CC mutant enzymes and the nucleic acids encoding them. The xyloglucanase of  
CC the invention can be used in the degradation of cellulose- and  
CC hemicellulose-containing biomass to produce ethanol. It can also be used  
CC in a detergent composition for treating fabric during a machine washing

CC cycle. The present sequence represents the Trichoderma reesei Family 74  
CC xyloglucanase.  
XX  
XX Sequence 838 AA;  
Query Match 31.7%; Score 1625.5; DB 7; Length 838;  
Best Local Similarity 41.7%; Pred.No.9.3e-78;  
Matches 350; Conservative 111; Mismatches 296; Indels 63; Gaps 23;  
QY 25 SFAVAALGLVLPALATASPAHAATTOPTYSNVAI--GGGFGVDGIYFNEGAPILYVRTD 83  
Db 4 SRVLAIVLGAV-----IAHAA-----FSWKVYKLGCGGGFGVGIIFHPRTKVAARFD 53  
QY 84 ICGMTRMDAANGRWIPLDLVW---GMNNGXGVVSIADPINTKMAAAGMTYNSWD 139  
Db 54 IGLVRLN--ADSSWPAVTGDIADNAGMHV---GIDAVALDPDDQKVVAAVGMTYNSWD 109  
QY 140 PNDGALIRSDGATMOTITPLPFLKGNMGRKMGRLAVDPENDNIIYFGAPSSGRLR 199  
Db 110 PGNAGIIRSDGATWSTFNLPPKVGNNPGRAGRLAVDPANSMIYFGAASGGLMK 169  
QY 200 STDSGATMGOMTFNPDPVGTIYIANPTDTTGYSODIQGVVWVAFDPKSSSLGQASKTIFGV 259  
Db 170 STDGVTFSKVSSTFATGTIIPDPSDNGNSDKQGLMWTFTDSTSTTGATSRIFVGT 229  
QY 260 ADP--NNPVWMSRDGATWQAVPGAPYGFIPHKGVPPVNVHLYIATSNTEGPGYSSGDV 318  
Db 230 ADNITASVYVSTNAGSTWGAIVGQPKGYPRHAKLQPAKALYLTYSDDGIPDGLGAV 289  
QY 319 WKPSTSGTWTRISPYPSYDTIANDYFGYGLTIDROHPNTINVAIOISWPDITIIIRSTD 378  
Db 290 WRDYDIAGTWDKIDITPVSGLD---YFGFGIGLIDLOKPGTLVVASINSWMPDQQLRSTD 346  
QY 379 GGATWTRIMWMTSYNPRSLRYVLDISAEPLTFEG--VQNPVPVS-----PKLGMDAMAI 433  
Db 347 SGTTPSPIMAMASYPLETYYSISTPKAPIKNFIIDVTSESSDGLIKRLGMEISLEI 406  
QY 434 DPFNSDRMLYGTGATLYATNDLTKWDSCGOIHIAIPWVKGLEETAVNDLISPGCAPLISA 493  
Db 407 DPTDNNHMLYGTGOMTIFGCHDLTNMDTRHNVLSQSLADIEBSYVDLNASRSGSLLLA 466  
QY 494 LGDLGGFTHA---DYTAVPSTIFTSPVFTTGSVDYAEINPSIIYRAGSFDPSSQPNDRH 550  
Db 467 VGDDNGFTFASRNDIGTSPQTWATPPTMATSTSVDAAGNSVSVAVGNTAGTQ----- 521  
QY 551 VAFSTDGKNNPQSGEPGVTGCTVAAADGSRFMAAGDQGPVYVAVFGNSMAASQ 610  
Db 522 VAISSDGGATWSIDVAADTSMNGGTVAVSADGDTIIMSTASG---VORSQFGSFAVS 578  
QY 611 GVPANAAQIRSDRVNPKTYALSNGTFFRSTDGGVTFPQVAAGLPSSGAVGVW--FHAVPG 668  
Db 579 SLPGAQVILASDKKTNVSIFYAGSGSTFYVSKDTGSSF---TRG-PKLGSKTIRDIANAHT 634  
QY 669 KEGDMLAASSGLYHSTNGSSWSAI--TGVSAVNVGFKSAPGSSYPFAVFGVGTIGVT 727  
Db 635 TAGTLYVSTDVGIFSTDSGTTFGQVSTALNTYQALGVGS--NLVAFGT--GRS 690  
QY 728 GA--YRSDCGTTLVINDOHOYGNWGAQITGDHANLRVYIIGTNGRGIVY--GDIGA 783  
Db 691 GARLYASGDSGASWTDIQSGQFGSIDSTKVGSGSTAGQVYVYGTNGRGVFAQGVGCG 750  
QY 784 PSGSPSPSVSPASPSLSPSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS 843  
Db 751 TGGT-----SSSTKQSSSTSSASSTLRSSVSTTRASTVTSRTSSAAGPTGS 801  
RESULT 9  
ID ADD424919 standard; protein; 555 AA.  
XX ADD424919;  
XX  
XX 15-JAN-2004 (first entry)  
DT

XX Xanthomonas campestris cellulase #1.  
 DE Directed genetic engineering; galactomannanase; reduced activity;  
 XX enhanced activity; xanthan gum production; suspension stability;  
 KM emulsion stability; temperature resistance; pseudolacticity; amylase;  
 KM cellulase; extracellular protease; intracellular protease;  
 KM glucose dehydrogenase; enzyme.  
 XX  
 OS Xanthomonas campestris.  
 XX  
 PN US2003036176-A1.  
 XX  
 XX 20-FEB-2003.  
 XX  
 XX 10-AUG-2001; 2001US-00927827.  
 PF  
 XX 28-MAR-2001; 2001US-0279493P.  
 PR  
 XX (BOWE/) BOWER S G.  
 PA (RAMS/) RAMSEIER T M.  
 PA  
 XX Bower SG, Ramseier TM;  
 PI  
 XX WPI; 2003-625389/59.  
 XX  
 DR N-PSDB; ADD24893.  
 DR  
 XX  
 PT New transformed cell or organism having reduced or enhanced activity of  
 PT at least one protein, useful for producing xanthan gum, which are useful  
 PT for providing formulations and properties.  
 PT  
 XX  
 XX Claim 1; SEQ ID NO 47; 135pp; English.  
 XX  
 CC The present invention relates to polypeptide and polynucleotide sequences  
 CC from Xanthomonas campestris which may be used for activity reduction or  
 CC enhancement using directed genetic engineering. A transformed cell or  
 CC organism having reduced or enhanced activity of at least one such protein  
 CC e.g. galactomannanase can be generated by disrupting the gene encoding  
 CC the protein. The activity of the protein is reduced by the presence of an  
 CC antisense nucleic acid sequence. The nucleic acid sequence of the gene  
 CC encoding the protein is a recombinant sequence having at least one  
 CC mutation as compared to the wild-type gene encoding the protein. The  
 CC transgenic cell or microorganism are useful for producing xanthan gum,  
 CC which are useful for providing formulations and properties, such as long-  
 CC term suspension and emulsion stability in alkaline, acid, and salt  
 CC solutions, temperature resistance, and pseudolacticity. The present  
 CC sequence represents an enzyme relating to the present invention.  
 CC  
 XX  
 XX Sequence 555 AA;  
 SQ  
 Query Match 16.5%; Score 848; DB 7; Length 555;  
 Best Local Similarity 36.5%; Pred. No. 7.3e-37;  
 Matches 193; Conservative 87; Mismatches 203; Indels 46; Gaps 14;  
 XX  
 QY 23 TASFAVAALGVLPATITASP----AHATTQPYTMSNAIGGGFVDIVNEGAPGIL 78  
 DB 46 TTGGAMLAIVLSL-LLRSTPSVRAAEPATSGPYKRSVAIGGGGVTVLHPHARGIA 104  
 QY 79 YRTDIDGMYRMDAANGRMIPLLDVGNNMGVNGVSIADPIINTKMAAVGMYTNSW 138  
 DB 105 YARDVVGAGAYRMDAQOQWTAALTDMIGADDMIMGIDAPVADADALYLAAGTYMH- 163  
 QY 139 DPNAGATIRSSDQATWQITPLPFLGGMPPRGNGERLAUVPNNNDNITIFGAPSKGIW 198  
 DB 164 RAGTAAVIRSRFRGRTFERADLPFLGQQLGRANGERLAUVPDHRGVLLGSRA-GLW 222  
 QY 199 RSTDSGATWSQMTNPPDVGYTIANPTDTGYQSDIGVYVAFDSSSSILGQASKTIIFVG 258  
 DB 223 RSDRGAMMAVYASFPDAALAGATARNHVRGQAV-GIAFVFPDAASGNGSPRTPIYVG 281  
 QY 259 VADBNPPVFWSRDGGATWQAVPGAPGTGFLPHKGVDPVNVHLYIATSNTPGYPYSSGDV 318  
 DB 282 VSTGQTSLYVSEBDAGRSWAPVAGPGRGLRPSHWAGSGDH-WYLSYGDQPGFDLMAGAL 340

QY 319 MKPSTVSGTWTRISPVSTDTANDYFGYGLTDRQHPNTIMVATQISMKWPTIIRSTD 378  
 DB 341 MKFTTPAQGRWREISPIQ-PASGDGFGMGAVAVDPQHPQVTLASTFRRRPRRDELYRSVD 399  
 QY 379 GGATWRIRMDWMSYPRKSLRYVLIDISAEPMLTFGVQPNPPVPSPKLGWDEAMALDPFNS 438  
 DB 400 GGGHWTPL-----LDAVAFDHSAAFP-TYHATPH-----WKG-ALALDPFDG 439  
 QY 439 DRMLYGTGATLYATNDLTWKDSGQJHIAP-----WVKGLETPAVNDLISPPSGAPL 490  
 DB 440 NHRLFTVGTGIMASRL-----QDPAPQRPLQWPFQRGLEETVPLDLSPMACAHL 492  
 QY 491 ISALGDLGGFTHADVTAVPSTIIFTSVFTTGSVDYAELNPSIIVRGS 539  
 DB 493 LSAIGDIDGFRHDDLDRV-QLGYAGPRLTNGSISIDAAGAPQWVRSRGT 540  
 RESULT 10  
 ABP73029  
 ID ABP73029 standard; protein; 1228 AA.  
 XX  
 XX ABP73029;  
 AC  
 XX 03-JUN-2003 (first entry)  
 DT  
 XX  
 XX Amino acid sequence of the GuXA polypeptide.  
 DE  
 XX GuXA, cellulase; glycoside hydrolase; enzyme; cellulose; sugar; biofuel;  
 KM detergent; pulp processing; paper processing; feed processing; textile.  
 KM  
 XX  
 XX Acidothermus cellulolyticus.  
 OS  
 XX  
 XX WO2003012109-A1.  
 PN  
 XX 13-FEB-2003.  
 XX  
 XX 28-JUL-2001; 2001WO-US023817.  
 PD  
 XX  
 XX 28-JUL-2001; 2001WO-US023817.  
 PF  
 XX  
 XX (MIDE ) MIDWEST RES INST.  
 PR  
 XX Ding S, Adney WS, Vanzant TB, Himmel ME, Decker SR;  
 PA  
 XX WPI; 2003-239526/23.  
 PI  
 XX N-PSDB; ABZ77634.  
 DR  
 XX  
 XX Novel thermal tolerant GuXA polypeptide derived from Acidothermus  
 PT cellulolyticus, useful for reducing cellulose in a starting material, and  
 PT for the conversion of biomass to biofuels and biofuel additives.  
 PT  
 XX  
 XX Claim 3; Page 19; 47pp; English.  
 XX  
 CC The present sequence represents a GuXA polypeptide. GuXA is thermostable  
 CC cellulase, and is a member of the glycoside hydrolase family of enzymes.  
 CC GuXA is useful for reducing cellulose in a starting material such as  
 CC agricultural biomass to sugars. This is useful in biofuel production.  
 CC GuXA is also useful in the conversion of biomass to biofuels and biofuel  
 CC additives, in detergent, pulp and paper processing, food and feed  
 CC processing, and in textile process. GuXA is also useful for raising  
 CC polyolonal and monoclonal antibodies that are useful in purifying GuXA,  
 CC or detecting GuXA polypeptide expression, as well as reagent tools for  
 CC characterizing the molecular actions of GuXA polypeptides  
 CC  
 XX  
 XX Sequence 1228 AA;  
 SQ  
 Query Match 16.0%; Score 823; DB 6; Length 1228;  
 Best Local Similarity 38.7%; Pred. No. 3.7e-35;  
 Matches 242; Conservative 44; Mismatches 189; Indels 150; Gaps 22;  
 QY 426 WDEAMALDPFNSDRMLYGTATLYATNDLTWKDSGQJHIAPWVKGLETPAVNDLISPP 485

```
Db 103 WMDRIALANGVNG-----GPGITTYLDAALSO-----QGGTTPREV---IEIYIYDL--RG 147
Qy 486 SGAPLISALGDLGFTHADVTAVPSTIFTSPPVTTGTSDVAELNPSIIVRASFPSSSQ 545
Db 148 RDCAALASNGELPA-TAAGLQTY-ETQYIDPLASILSNKYSLSRLVTII-----EPDSL 200
Qy 546 PN---DRHVAISTDGGKQKMFQSGE-----PGCVTTGGTVLAASADGSRFVMAFGDGG 593
Db 201 PNAVTNMSIQACATAVPYEEQIEYALTKLHAIP-----NVYIYMDAAHSGMTL-GWPN 252
Qy 594 QPVVAVAFGKSMASQGV-----PANAQIRSRVNPKTFYALS 632
Db 253 NASGVQEVQKVLNLSIGVNGIDGVTYNTANTPLKEPMTXTOQGGQPVESANTFYQNN 312
Qy 633 NGTFYRSTGCVTFOPVAAAGLPSGAVGVMFHVPKEGEDLWLAASSGLYHSTNGSSWS 692
Db 313 PDIDBADVAVDLYSRLVAAAGFPSS--IGMLIDTL-----RNGMGGN 352
Qy 693 AITGVSSAVNV-----GFGKAPGSSYPAVF----- 718
Db 353 EPTGPSTATDVNTFVNQSKIDLRQHRGLMCNQGAGLQPPQASPTDFPNAHLDAYVMIK 412
Qy 719 VVGITIGVTGA-----YRSDD-CGTTWVLINDQHOYGNMQA-----ITGDHANLRRV 766
Db 413 PPGESDGTSAASDPTTGKKSDPMCDPTTYT-----SYGVLTNALPNSPIAGQWPPAQFD 466
Qy 767 YIGTNGRGIVYGDIG-----APSGSPSPSVSPASPSLSPSPSSSPSPSP----- 814
Db 467 QLVANARAVPRTSTSSPPPPSPSASPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS 526
Qy 815 ---SPSSSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 871
Db 527 PSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 586
Qy 872 GVKQYQKNDAPGNOIKPGLQVNTGSSSYDLSTVYTRVWFTRDGSSSTLYNCDMAA 931
Db 587 GLKQYQKNDAPGNOIKPGLQVNTGSSSYDLSTVYTRVWFTRDGSSSTLYNCDMAA 646
Qy 932 IGGGNIRASFGSVNPATPTADTYLQ 956
Db 647 MCGGNIRASFGSVNPATPTADTYLQ 671

RESULT 11
ABP73022
ID ABP73022 standard; protein; 762 AA.
XX
AC ABP73022;
XX
DT 03-JUN-2003 (first entry)
XX
DE Amino acid sequence of the Mana polypeptide.
XX
KW Mana; mannanase A; glycoside hydrolase; enzyme; hemicellulose; sugar;
XX food; feed; paper pulp; biofuel; mannanase.
XX
OS Acidothermus cellulolyticus.
XX
PN WO2003012110-A1.
XX
PD 13-FEB-2003.
XX
PF 28-JUL-2001; 2001WO-US023819.
XX
PR 28-JUL-2001; 2001WO-US023819.
XX
PA (MIDE ) MIDWEST RES INST.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX
WP1; 2003-248182/24.
XX
DR N-PSDB; AB277633.
XX
```

```
PT Novel thermal tolerant mannanase A polypeptide derived from Acidothermus
PT cellulolyticus, useful for reducing hemicellulose in a starting material,
PT for processing of food, and as bulking agents in food stuffs.
XX
XX Claim 3; Page 18-19; 46pp; English.
CC The present sequence represents Mana, a thermostable mannanase A
CC polypeptide derived from Acidothermus cellulolyticus. Mana is a member of
CC the glycoside hydrolase family of enzymes. Mana is useful for reducing
CC hemicellulose in a starting material to simpler carbohydrate units, and
CC ultimately to sugars which are useful in the food, feed, paper pulp, and
CC biofuels industries. It is useful for the processing of food and in food
CC stuffs as bulking agents, and for the degradation of mannanase. Mana is
CC also useful to raise polyclonal and monoclonal antibodies that are useful
CC in purifying Mana, or detecting Mana polypeptide expression, and as well
CC as reagent tools for characterizing the molecular actions of Mana
CC polypeptides
XX
SQ Sequence 762 AA;
XX
Query Match 13.7%; Score 702; DB 6; Length 762;
Best Local Similarity 35.0%; Pred. No. 5,4e-29;
Matches 214; Conservative 53; Mismatches 148; Indels 196; Gaps 27;
Qy 472 GLEETAVNDLISPPSGAPLISALGDLGFTHADVTAVPSTIFTS-----PVFTTG 521
Db 2 GLVRRPARAFVAAAGT-AVAAAATLGSITMPATAPAFVAAAGQFVNLGLPYRYGG 60
Qy 522 TSVDYAE-----LNPSII-----VAAQSPSPSQPNDRHVAISTGGKN 560
Db 61 TNNYIYSOSHADVDDVLAKAQMNLIVLRWGFIDIGSLD-GSVP-----TIGNNKN 112
Qy 561 --WFGSBEPCGVTTG-----GTVAASADGSRFVMAFGDQPVVAVAVGF 602
Db 113 GPFYQWDP---STGAPAYNDGPTGLQGLDYALASAAHGLRIY-----VVL 156
Qy 603 GNSWAAS-----QGVPAQAQIRSDRVNPRTFYALNGTFYR---STDGVTF--- 646
Db 157 TNDWKEFGMDQDYDKWYGLPYHDFNYTD---PRTQAQYKQWVHLLNRRVNSIGVTYKND 213
Qy 647 -----QPVAAQ---LPSSGA-----VGMFPAVKGKEDMLA-ASSGLYH 663
Db 214 PTFIAMELANEPRCVCVSGTLPTSGTCTOATIVMWVQMSAYVKSIDPNHNVSVGDGCFYI 273
Qy 684 STWGSSW---SAITGVSS-----AVNVGFGKAPGSSYPVAVVGTIGVTGAYRSDC 725
Db 274 GSTQSGGMPYNDPDSGDVNNALLRVKNIIDPGTY---HLIPNVY-----GQNAW 319
Qy 736 GTTWV---LIN-----DDQHQYGNWGAITGDHANLRRVYIGTNGR- 773
Db 320 GTQWIKQHIANAALIGKPTLIEFGWOTPRDGVYQTVQTVR-----TNGEA 367
Qy 774 GIVYDIDGAPSGSPSPV-----SPASPSLSPSPSSPSPSPSPSPSPS 825
Db 368 GNMFMWLAGVNCQPYPNYDGFVNVYPSSTAYTLASEALALISTGTPSPSPSSPS 427
Qy 826 PSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 885
Db 428 PSPSPSPSPSP-----SASSPSPSPSPSVSGVQYQYNNNSAPG 471
Qy 886 DNOIKPGLQVNVNTGSSSYDLSTVYTRVWFTRDGSSSTLYNCDMAAIGCNIRASFGSVN 945
Db 472 DNOIKPGLQVNVNTGSSSYDLSTVYTRVWFTRDGSSSTLYNCDMAAMGCGNIRASFGSVN 531
Qy 946 PATPTADTYLQ 956
Db 532 PATPTADTYLQ 542

RESULT 12
ABP73018
ID ABP73018 standard; peptide; 88 AA.
XX
```

AC ABB73018;  
XX 03-JUN-2003 (first entry)  
XX  
XX Amino acid sequence of AvilIII carbohydrate binding domain type II.  
DE  
XX Avicelase; AvilIII; glycoside hydrolase; enzyme; cellulase; biofuel;  
KM detergent; pulp processing; paper processing; feed processing; textile;  
KM cellulose.  
XX  
OS Acidothermus cellulolyticus.  
XX  
XX WO2003012090-A2.  
XX  
XX 13-FEB-2003.  
XX  
XX 28-JUL-2001; 2001WO-US023818.  
XX  
XX 28-JUL-2001; 2001WO-US023818.  
XX  
XX (MIDE ) MIDWEST RES INST.  
XX  
XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX WPI; 2003-248177/24.  
XX  
XX  
XX New thermostable AvilIII peptide from Acidothermus cellulolyticus, useful  
PT for degradation of cellulose or in generating anti-AvilIII antibodies for  
PT purifying recombinant AvilIII polypeptides from genetically engineered  
PT host cells.  
XX  
XX Claim 6; Page 8; 44pp; English.  
XX  
CC The present sequence is derived from a thermostable avicelase, designated  
CC AvilIII. AvilIII is a member of the glycoside hydrolase family of enzymes,  
CC and is a cellulase. AvilIII is useful in the conversion of biomass to  
CC biofuels and biofuel additives. It may be useful in the production of  
CC detergents, pulp and paper processing, food and feed processing and in  
CC textile processes. The thermostable AvilIII peptide is useful in the  
CC degradation of cellulose, and in generating specific anti-AvilIII  
CC antibodies that are useful in purifying recombinant AvilIII polypeptides  
CC from genetically engineered host cells, in detecting AvilIII polypeptide  
CC expression, as well as a reagent tool for characterizing the molecular  
CC actions of the polypeptide. The AvilIII polynucleotide is useful as a  
CC source of probes or primers in various diagnostic assays  
XX  
XX Sequence 88 AA;  
SQ  
Query Match 9.1%; Score 468; DB 6; Length 88;  
Best Local Similarity 100.0%; Pred. No. 1.1e-17;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 869 VSGGVKQYKKNDSAPGDNQIKFGLQVNTGSSVDLSTVTRVYWFTRDGGSSTLVYVNC 928  
DB 1 VSGGVKQYKKNDSAPGDNQIKFGLQVNTGSSVDLSTVTRVYWFTRDGGSSTLVYVNC 60  
QY 929 WAAIGCGNIRASFGSVNPAFTADTYLQ 956  
DB 61 WAAIGCGNIRASFGSVNPAFTADTYLQ 88  
RESULT 13  
ABB73017  
ID ABB73017 standard; peptide; 89 AA.  
XX  
XX ABB73017;  
XX  
XX 03-JUN-2003 (first entry)  
XX  
XX Amino acid sequence of AvilIII carbohydrate binding domain type III.  
DE  
XX Avicelase; AvilIII; glycoside hydrolase; enzyme; cellulase; biofuel;  
KM detergent; pulp processing; paper processing; feed processing; textile;  
KM

KM cellulose.  
XX  
XX Acidothermus cellulolyticus.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 89 /note= "unspecified residue"  
FT  
XX  
XX WO2003012090-A2.  
XX  
XX 13-FEB-2003.  
XX  
XX 28-JUL-2001; 2001WO-US023818.  
XX  
XX 28-JUL-2001; 2001WO-US023818.  
XX  
XX (MIDE ) MIDWEST RES INST.  
XX  
XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
XX WPI; 2003-248177/24.  
XX  
XX  
XX New thermostable AvilIII peptide from Acidothermus cellulolyticus, useful  
PT for degradation of cellulose or in generating anti-AvilIII antibodies for  
PT purifying recombinant AvilIII polypeptides from genetically engineered  
PT host cells.  
XX  
XX Claim 6; Page 8; 44pp; English.  
XX  
CC The present sequence is derived from a thermostable avicelase, designated  
CC AvilIII. AvilIII is a member of the glycoside hydrolase family of enzymes,  
CC and is a cellulase. AvilIII is useful in the conversion of biomass to  
CC biofuels and biofuel additives. It may be useful in the production of  
CC detergents, pulp and paper processing, food and feed processing and in  
CC textile processes. The thermostable AvilIII peptide is useful in the  
CC degradation of cellulose, and in generating specific anti-AvilIII  
CC antibodies that are useful in purifying recombinant AvilIII polypeptides  
CC from genetically engineered host cells, in detecting AvilIII polypeptide  
CC expression, as well as a reagent tool for characterizing the molecular  
CC actions of the polypeptide. The AvilIII polynucleotide is useful as a  
CC source of probes or primers in various diagnostic assays  
XX  
XX Sequence 89 AA;  
SQ  
Query Match 9.1%; Score 468; DB 6; Length 89;  
Best Local Similarity 100.0%; Pred. No. 1.1e-17;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 869 VSGGVKQYKKNDSAPGDNQIKFGLQVNTGSSVDLSTVTRVYWFTRDGGSSTLVYVNC 928  
DB 1 VSGGVKQYKKNDSAPGDNQIKFGLQVNTGSSVDLSTVTRVYWFTRDGGSSTLVYVNC 60  
QY 929 WAAIGCGNIRASFGSVNPAFTADTYLQ 956  
DB 61 WAAIGCGNIRASFGSVNPAFTADTYLQ 88  
RESULT 14  
ADD22925  
ID ADD22925 standard; protein; 88 AA.  
XX  
XX ADD22925;  
XX  
XX 15-JUN-2004 (first entry)  
XX  
XX Acidothermus cellulolyticus avicelase AvilIII CBD III #2.  
DE  
XX enzyme; AvilIII; cellulose reduction; agricultural biomass;  
KM municipal solid waste; glycoside hydrolase; avicelase.  
XX  
XX Acidothermus cellulolyticus.  
OS  
XX US2003108988-A1.  
PN

XX 12-JUN-2003.  
PD 18-OCT-2002; 2002US-00155400.  
XX 28-JUL-2001; 2001US-00917376.  
XX (DING/) DING S.  
PA (ADNEY/) ADNEY W S.  
PA (VINZ/) VINZANT T B.  
PA (HIMM/) HIMMEL M E.  
XX Ding S, Adney WS, Vinzant TB, Himmel ME;  
PI WPI; 2003-810853/76.  
XX New isolated thermal tolerant avicelase polynucleotide useful for  
PT detection of a polynucleotide encoding AvIII and for reducing cellulose  
PT in a starting material, e.g. municipal solid waste.  
XX Claim 16; SEQ ID NO 5; 29pp; English.  
XX The invention relates to an isolated polynucleotide molecule encoding a  
CC thermostable AvIII polypeptide. The polynucleotide is useful for  
CC detection of a polynucleotide encoding AvIII. The polynucleotide is  
CC useful for reducing cellulose in a starting material which involves  
CC administering to the starting material, e.g. agricultural biomass or  
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The  
CC method further comprises administering a second polypeptide molecule  
CC chosen from the glycoside hydrolase family of proteins. The present  
CC sequence represents the amino acid sequence of Acidothermus  
CC cellulolyticus avicelase AvIII CBD III #2.  
XX  
SQ Sequence 88 AA:

Query Match 9.0%; Score 462; DB 7; Length 88;  
Best Local Similarity 97.7%; Pred. No. 2.3e-17;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 869 VSGGVQYKNNDSAPGDNQIKRGLQVNTGSSVDLSTVYRYWPTRDGSSSTLYNCD 928  
DB 1 VSGGVQYKNNDSAPGDNQIKRGLQVNTGSSVDLSTVYRYWPTRDGSSSTLYNCD 60

QY 929 WAIIGGNTIRASFGSVNPAFTADTYLQ 956  
DB 61 WAIIGGNTIRASFGSVNPAFTADTYLQ 88

RESULT 15  
ADD22924  
ID ADD22924 standard; protein; 89 AA.

XX ADD22924;

DT 15-JAN-2004 (first entry)

DE Acidothermus cellulolyticus avicelase AvIII CBD III #1.

KW enzyme; AvIII; cellulose reduction; agricultural biomass;  
municipal solid waste; glycoside hydrolase; avicelase.

OS Acidothermus cellulolyticus.

XX Key Location/Qualifiers

FT Misc-difference 89

XX US2003108988-A1.

XX 12-JUN-2003.

XX 18-OCT-2002; 2002US-00155400.

PR 28-JUL-2001; 2001US-00917376.

XX (DING/) DING S.

PA (ADNEY/) ADNEY W S.

PA (VINZ/) VINZANT T B.

PA (HIMM/) HIMMEL M E.

XX Ding S, Adney WS, Vinzant TB, Himmel ME;

PI WPI; 2003-810853/76.

XX New isolated thermal tolerant avicelase polynucleotide useful for

PT detection of a polynucleotide encoding AvIII and for reducing cellulose

PT in a starting material, e.g. municipal solid waste.

XX Claim 16; SEQ ID NO 4; 29pp; English.

XX The invention relates to an isolated polynucleotide molecule encoding a

CC thermostable AvIII polypeptide. The polynucleotide is useful for

CC detection of a polynucleotide encoding AvIII. The polynucleotide is

CC useful for reducing cellulose in a starting material which involves

CC administering to the starting material, e.g. agricultural biomass or

CC municipal solid waste, a polypeptide molecule of the polynucleotide. The

CC method further comprises administering a second polypeptide molecule

CC chosen from the glycoside hydrolase family of proteins. The present

CC sequence represents the amino acid sequence of Acidothermus

CC cellulolyticus avicelase AvIII CBD III #1.

XX  
SQ Sequence 89 AA:

Query Match 9.0%; Score 462; DB 7; Length 89;  
Best Local Similarity 97.7%; Pred. No. 2.4e-17;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 869 VSGGVQYKNNDSAPGDNQIKRGLQVNTGSSVDLSTVYRYWPTRDGSSSTLYNCD 928  
DB 1 VSGGVQYKNNDSAPGDNQIKRGLQVNTGSSVDLSTVYRYWPTRDGSSSTLYNCD 60

QY 929 WAIIGGNTIRASFGSVNPAFTADTYLQ 956  
DB 61 WAIIGGNTIRASFGSVNPAFTADTYLQ 88

Search completed: May 14, 2004, 09:16:26

Job time : 70 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2004, 09:14:08 ; Search time 27 Seconds  
(without alignments)  
3409.455 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MDRSENRLTMRSRRLVSL.....RASFGSNPATPTADTYLQX 957

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2436	47.4	890	2 T35237	probably secreted
2	2037	39.7	839	2 D97013	secreted
3	1740	33.9	856	2 T00349	Avicelase III - As
4	1161	22.6	707	2 F72393	hypothetical prote
5	431.5	8.4	1779	2 T31085	xylanase - Caldice
6	398.5	7.8	1711	2 T31337	1,4-beta-glucanase
7	397.5	7.7	1742	2 T17120	cellulase (EC 3.2.
8	372.5	7.3	473	2 S50755	hypothetical prote
9	346.5	6.7	901	2 A49227	slalidase - Actino
10	340.5	6.6	1331	2 A48954	mannan endo-1,4-be
11	308.5	6.0	611	2 S76211	hypothetical prote
12	301.5	5.9	1749	2 S75138	hypothetical prote
13	293	5.7	279	2 T10361	hypothetical prote
14	291	5.7	351	2 S50754	hypothetical prote
15	290.5	5.7	2232	2 T34434	hypothetical prote
16	290	5.6	1039	2 S02711	cellulase (EC 3.2.
17	288.5	5.6	1032	2 T34433	hypothetical prote
18	283.5	5.5	474	2 S15921	protein TPX-VT3 -
19	262	5.1	2468	2 A63412	hypothetical prote
20	261.5	5.1	1854	2 S36859	clpA protein - Clo
21	254.5	5.0	913	2 S20590	exo-alpha-sialidas
22	246	4.8	915	2 A43802	cellulase (EC 3.2.
23	245	4.8	879	2 A47704	endoglucanase I (E
24	244	4.8	1664	2 T18262	S-layer protein -
25	243.5	4.7	552	2 T08148	proline-rich myros
26	243	4.7	288	2 T17737	proline-rich prote
27	235.5	4.6	2124	2 A28452	proteoglycan core
28	229.5	4.5	496	2 T17908	proline/lysine-ric
29	227.5	4.4	2271	2 F90073	hypothetical prote

30	222.5	4.3	360	2 S12850	protein TPX - Ther
31	222.5	4.3	3570	2 T45025	mucin MUC5B, trach
32	221	4.3	2132	1 A55182	aggreacan precursor
33	219	4.3	4776	2 E95206	cell wall surface
34	218.5	4.3	3164	1 WMBBHE	U36 protein - hum
35	217	4.2	241	2 AC2284	hypothetical prote
36	216	4.2	383	2 T46707	hypothetical prote
37	216	4.2	449	2 S16748	proteophosphoglyca
38	215.5	4.2	13055	2 T16580	proline-rich prote
39	215	4.2	339	2 T17636	hypothetical prote
40	215	4.2	13288	2 T03099	proline-rich prote
41	214	4.2	300	2 J02320	mucin, submaxillar
42	214	4.2	532	2 B35621	hydroxyproline-ric
43	211.5	4.1	1275	2 T33369	spore germination
44	211	4.1	602	2 AD2067	hypothetical prote
45	210	4.1	549	2 T17525	proline-rich prote

ALIGNMENTS

RESULT 1	
T35237	
probable secreted cellulase - Streptomyces coelicolor	
C/Species: Streptomyces coelicolor	
C/Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999	
C/Accession: T35237	
R/Seeger, K.J.; Harris, D.; Parkhill, J.; Barrrell, B.G.; Rajandream, M.A.	
submitted to the EMBL Data Library, September 1998	
A/Reference number: Z21572	
A/Accession: T35237	
A/Status: preliminary; translated from GB/EMBL/DBJ	
A/Molecule type: DNA	
A/Residues: 1-890 <SBE>	
A/Cross-references: EMBL:AL031515, PIDN:CMA20642.1, GSPDB:GN00070, SCODEB:SC5C7.30C	
A/Experimental source: strain A3(2)	
C/Genetics:	
A/Gene: SCODEB:SC5C7.30C	
Query Match	47.4%; Score 2436; DB 2; Length 890;
Best Local Similarity	48.6%; Pred. No. 1.7e-102;
Matches 475; Conservative 129; Mismatches 248; Indels 126; Gaps 19;	
QY	11 MRSRRVLSLAATASPAVAALGVLPATASPA--HAATTPYTWNSNAIGGGGFVDGI 68
DB	1 MRRTRITVTLAALAGLA---GSPRAASAEAPRAVAADSYTWGNARIDGGFVPGI 57
QY	69 VFNEGAPGLVYRTDIGMYRMDANGRWIPLLDWVGNNMGVNSIADPINTNKW 128
DB	58 VFRRTEDLAVARTDIGAVRWQEEHWTPLLDHVQMDMGHTGVVALASDAVDPDXY 117
QY	129 AAVGMYTNSWDPDGAILSSDQATWQITPLFPKLGNNPGKMGKRLAVDNNINLY 188
DB	118 AAVGYTNDWDPDNGVNLRSADRGSWEKADLPFKLGNNPGKMGKRLAVDHDNDVLY 177
QY	189 FGAPSGKGLMRSDGATWSQMTNPPDVGTYINPTDTTGYOSPDIGVVAVAFDKSS--SS 247
DB	178 LGAPSGHGLMRSDAGVTSVTAFAFPGNIADPNDTSGLASDNGGIVTWTDSESTGG 237
QY	248 LQGSRTIIFVGVADPNNPVFWSRSDGATWQAVGAPTFGLPHKGVDPVNVHLYIATSNT 307
DB	238 AGTARTLVYGVADKENVAYRSTDAGTWRLAGQPGTYLAHNGVLAENGYLYLAVSDT 297
QY	308 GGPYDSSGDVWKFVSGTWRISVPVSTDTANDYFGYSGLTIDRQHPNTINAVATQISW 367
DB	298 GGPYDGGKRLVYATATGTWTDISPAAEADT---YGFSGLTVDRORPETVAVATVSSW 354
QY	368 WPTTITFRSDGATWTRIMDWSYPRSLRYVLDISAEWMLFRGVQPNPVPSPKLGW 427
DB	355 WPTQIFRSTDSGATWSQAWSYSPDRENRMTWDSVSWLWGAAPAPPEQTPKLGW 414
QY	428 DEAMADPFNSDMLTGTGATLYATNDLTKM--DSGQIHTAPWVKGLEAFVAVNDLISPPS 486

Db 415 TEALRIDPDSRMMYTGATVYGTENTLNMDDEGTFAPVEPMVRLGETAVNDLASPPS 474

Qy 487 GAPLISALGDIAGFTHADVTAVPSTIFTSPTVETGTSDVYAEINLSIIVRAGSPSSOP 546

Db 475 GAPLISALGDIAGFTHADVTAVPSTIFTSPTVETGTSDVYAEINLSIIVRAGSPSSOP 546

Qy 547 NDHVAEPTDGGKMPQSGEPGVTTGTGTAASADGSRFVWAPGDPQGVYAVGFGNSW 606

Db 533 ---HIAFTDNGAMWFGSTDPGSGVAGGTAAGADGSRFVWAPGDPQGVYAVGFGNSW 587

Qy 607 AASGVANPAQIRSDRVNPKTFYALNSGTFRSTGTGTFQGVPAV-GLPSSAGVGMFLA 665

Db 588 QASGFLPGALVSDRVNPAVTFYKSRFVSTGTGATFTASATGTPAGD--GVRFYA 645

Qy 666 VPKGEGDLMLAAS-----GLYHSTNGSSWGAITGVSSAVNVGFKGAPSGSYPAVFW 720

Db 646 LPGEHGDVWLKGADGPGYGLMHSIDGGGTFTRLRGVDAADLVYKGAAPGASGYQLFPTS 705

Qy 721 GTTGGVTVAYRSDGCTTGVTLINDQHOYGNMGOAITGDHANLRRVYIGTNGRGIVYGD 780

Db 706 AEIGGVGIFRSTAGATWTRVNDADHGMGTGAITGDPVRYGRVYATNGRGVYGD- 764

Qy 781 GGAEGSPSPSVSPSASLSLSPSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSP 840

Db 765 -----TSPTGCGTDPGPGPDPPTP----- 782

Qy 841 SASPSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSP 900

Db 783 -----TGACEVITYTNTQMPGFGFQ--ADVRLNTGT 811

Qy 901 S-----SYDLS-----TVTRVWFTF--DGGSSTLVYNCW-----AAIGCGNIRA--- 939

Db 812 SAMNGWGLDWFPGQGEVTRMMNAHTQAGTSVTAARVNMNAGVAPGASVGFSGFTSRSG 871

Qy 940 -----SFGSVNPAETPA 951

Db 872 TNAEPGFAVAGRACPTA 889

RESULT 2

D97013

probably secreted sialidase, several ASP-boxes and dockerin domain [imported] - Clostrid

C/Species: Clostridium acetobutylicum

C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001

C/Accession: D97013

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A;Reference number: A96900; MUID:21359325; PMID:21359325

A;Accession: D97013

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-839 <KUD>

A;Cross-references: GB:AF001437; PIDN:AAK78895.1; PID:GL5023820; GSPDB:GN00168

A;Experimental source: Clostridium acetobutylicum ATCC824

C;Genetics:

A;Gene: CAC0919

Query Match 39.7%; Score 2037; DB 2; Length 839;

Best Local Similarity 49.5%; Pred. No. 1.4e-84;

Matches 392; Conservative 118; Mismatches 252; Indels 30; Gaps 16;

Qy 13 SRLVSLIAA-----TASPAVAALGVLEPIATSPAHAAAT--TQPYTWSNVAIGGGFVD 66

Db 2 NKRIYSWVAGLSIIFTTGF-----VTHISAANKAAASVQSGYKMNNAKIGAGGYVP 54

Qy 67 GIVEHEGAPGILYVRDIDGMYRMDAANGRWIPLDVGWMMNMGVGVSIADAPINTNK 126

Db 55 AVIFKTEKDIYIATDWMGAGARMDKANKWIPITD--GFSWMTMGCSISADPIDTNR 112

Qy 127 VMAAVGMYTNSWMDPDGAILRSSDQATWQITPLPKLGGNMPGRGMRGLAVDPNNNDI 186

Db 113 VYIAAGLYNMDODENAYIISODKGNWTKRYOLPRVGVGMNMRNGERLQIDPNDKI 172

Qy 187 LYFGAPSGKGLMRSTDSGATWSQMTNPPDVGYIANPTDTGYQSDIQGVVWAFDKSS 246

Db 173 LYFGAPSGKGLMRSTDSGATWSQMTNPPDVGYIANPTDTGYQSDIQGVVWAFDKSS 246

Qy 247 SLGQAKTTFVGVAD--PNNPFWSRDQATWQVAPGAPGFIPIHKGVFDPVNHVLIATS 305

Db 231 TKQSPQWTVVGAADDKTGNNIYVINDGKTWSAVKQPKGYLPHHGLI--ASDGLYISYS 289

Qy 306 NNGPFGSSGQVWKSYSVSTWTRISPVSTPDANDYFVSGTLTDROHPNTIMVATQ 365

Db 290 NNGPFGSSGQVWKSYSVSTWTRISPVSTPDANDYFVSGTLTDROHPNTIMVATQ 365

Qy 366 SWMPDITFSTGATWTRIMDTSYPNRSLRYVLDISAPMLTFG--VQPNPVPSPKL 424

Db 347 RMPDDEITRSTAGKTWRIWNGYPRNLTGLNLDYSAPMLDCKTGVTPEPDLVYL 406

Qy 425 GNMDMAIDPFNSDRLYGTGATLYATNDLYKWSGGQIHLAPWKLEETAVNDLISP 484

Db 407 GNMWGLDIDPFNSDRLYGTGATLYATNDLYKWSGGQIHLAPWKLEETAVNDLISP 484

Qy 485 PSGAPLISALGDIAGFTHADVTAVPSTIFTSPTVETGTSDVYAEINLSIIVRAGSPSS 544

Db 467 GNMWGLDIDPFNSDRLYGTGATLYATNDLYKWSGGQIHLAPWKLEETAVNDLISP 484

Qy 545 QPNDRVAVFSTGDKWFO--GSEPGVTTGTVAAASADGSRFVWAPGDPQGVYAVGFG 603

Db 527 NQDDKCCGISYDGGKMFSAKSNISGVYKAGTYAAGDAKTIWSP--EKGANAISTDNG 585

Qy 604 NSWAAQGVAPANAQIRSDRVNPKTFYALNSGTFRSTGTGTFQGVPAV-GLPSSAGVGMFLA 665

Db 586 NKTTPSGGLPQAGKAVSDRVNPKTFYALNSGTFRSTGTGTFQGVPAV-GLPSSAGVGMFLA 665

Qy 663 FHAVPKGEGDLMLA--ASSGLHSTNGSSWGAITGVSSAVNVGFKGAPSGSYPAVFW 720

Db 644 FKTIVGHEGDIWLAGKDIHSTDSGATFTKYSGVADSVTLVIGKSKRTDGDGPATIMDA 703

Qy 722 TGGVTVAYRSDGCTTGVTLINDQHOYGNMGOAITGDHANLRRVYIGTNGRGIVYGD 780

Db 704 TIDGTAGIFRSDGATWTRVNDADHGMGTGAITGDPVRYGRVYATNGRGVYGD- 764

Qy 782 GAPSGSPSPSVS 793

Db 764 GS-QPPTPESVT 774

RESULT 3

T00349

Avicelase III - Aspergillus aculeatus

C/Species: Aspergillus aculeatus

C/Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 16-Jul-1999

C/Accession: T00349

R;Arat, M.; Takada, G.; Kawaguchi, T.; Sumitani, J.

A;Title: The EMBL Data Library, June 1998

A;Description: Avicelase III from Aspergillus aculeatus.

A;Reference number: Z14141

A;Accession: T00349

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-856 <ARA>

A;Cross-references: EMBL:AB015511; NID:dl199887; PID:dl029971

C;Genetics:

A;Gene: avIII

C;Superfamily: fungal cellulose-binding domain homology

F;823-854/Domain: fungal cellulose-binding domain homology <PCB>

Query Match 33.9%; Score 1740; DB 2; Length 856;

Best Local Similarity 43.5%; Pred. No. 3.3e-71;

Matches 366; Conservative 139; Mismatches 237; Indels 40; Gaps 18;

Qy 30 AALGVPIAIIATSPAHAAATQPYTWSNVAI--GGGGFVDGIVFNGAGIILVTRDIGMY 88

4 SSLALLCALGLKADAAASQAYTWKAVNTVGGGGFTGCIYVNPBSAKGVAVARTIDGAY 63  
89 RMDAANGRMIDPLDMVG---WNNMGYNVSIADPINTNKWAAVGMTNSWDPNDGAI 145  
64 RLN-SDDTWTFELMDVWGNIDTWHDW---GIDALATPDVTDTRVYVAVGMTNEMDPRVSGI 119  
146 LRSDDGATWQITPLPFLGGMGRGGERLAUPNNNDIILYFGAPSGKGLMRSTDGA 205  
120 LRSTDOGTWETETKLPFYKVGGMGRGERLAUPNNKSIILYFARSGHGLMKSTDYGA 179  
206 TWSQNTNPDVGTIYANPTDTGTGYSIDIGYVWVAFDSSSSSLGQASKTIIFGVADPNNP 265  
180 TWSVNTSTMTGTTFODSSST--YTSDEPGIANTFDDTSSSGSATPRIFGVADAGS 237  
266 VFWSRDGGATWQAVPGAPT-GFIPIKGVFPDPVNHVLYATNSNGPYDSSGDVWKFVST 324  
238 VFKESEDAGATWAVSGEPQYGLPHKGLVLSPEKTLIYSYANGAPYDGTNGTVHKYNT 297  
325 SGWTRISPVSTDTANDYFGYSGLTIRQHPNTIMVATQISWMPDTIIFRSTDDGATWT 384  
298 SGVWTDISP---TSLASTYYGGYGLVDLQVPGTLMVVALNCWMPDELIIFRSTDSGATWS 354  
385 RIMWMTSYPNLSRLVLDISAEPMLTFGVQPNP-PVPEPKLGMDDEAMAIDPNSDRMLY 443  
355 PIMWMNGPISINYYISYDLSNAPWIDTTSDDQFV---RVGMVVEALAIIDPDSNHWLY 411  
444 GTGATLYATNDLTKMDSGGQIHIAIPMKLEETAVNDLISPPSGAPLISALDGLGFTTA 503  
412 GTGLTVGHHLTNMDSKHNTVKSIAVGIEMAVLGLITPPGAPLISAVDDDGFTYS 471  
504 DVTAVPSTIFTSPVFTGTSDVYAEALNPSIIVRAGSPDSSQPNDRHVAFTDGGKNMFO 563  
472 DLDAAPNQAYHTPTGTNGIDYAGNKRPSNIYVSGASD--DYPT---LALSNGFTWYA 526  
564 GSEEGCVTGTGTVAAASADGRFVWAPGDPQPVYVAVFGNSMAASQGVPAQAQIRSDYV 623  
527 DYAASTSTGTAVVLSADGDVTVLMSSTGALVSKSG---TLTAVSSLPISGAVIASDKS 583  
624 NPKTFYVLSNGTFRSTDGTGVTPOVAAGLPSSGAVGVAFHVPKGGGLMTAAASGLVH 683  
584 DNIVFYGSGAIIYVSKNTATSTFKTVS-LGSSSTVNAI-RAHPSIAGDVMASTDGKLVH 641  
684 STNGSSSSWAI-TGVSSAVNVGFGKAPGSSSYPAVFVGTIGCVGAYRSDCGTMTVL 742  
642 STDVGSFTTQIGSGVTAGMSPFGKASSTGSYVVIYGFETIDGAAGLFRSBDAGTMQYI 701  
743 NDDHQYGNMGA--ITGDHANLARVYIGTNGRGIYVDIGAPSG-----SPSPSV 792  
702 SDASHGFSG-GSANVNVNDLQTYGRVFRGHERFGHILROSQREPAGRHGDGDGDTTSKT 760  
793 SPSPSPSLSPSPSSSP 852  
761 STVSTLTKTTTSSASTTSSSTTVKTTSSSTTSKASTTYYTKTTTSTTSSGTTATA 820  
853 SS 854  
821 SA 822

RESULT 4  
F72393  
hypothetical protein TM0305 - Thermotoga maritima (strain MSB8)  
C/Species: Thermotoga maritima  
C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C/Accession: F72393  
R/Melison, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
Nature 399, 323-329, 1999  
A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A/Reference number: A72200; MUID:99287316; PMID:10360571  
A/Accession: F72393  
A/Status: preliminary

A/Molecule type: DNA  
A/Residues: 1-707 <ARN>  
A/Cross-References: GB:AE001712; GB:AE000512; NID:g4980799; PIDN:AAD35393.1; PID:g498080;  
A/Experimental source: strain MSB8  
C/Genetics:  
A/Genes: TM0305

Query Match 22.6%; Score 1161; DB 2; Length 707;  
Best Local Similarity 35.2%; Pred. No. 2,9e-45;  
Matches 264; Conservative 123; Mismatches 275; Indels 89; Gaps 26;

52 YTWNSVAIAGGGGVVDGIVPNEGAGTLYVTRTDIGWYRWMDAANGRMIPLDVWGMNMG 111  
21 FEKSVYEINGGGVPGGILTHPASPGLLYATDVGLRWBETKMKQLPDLRQSDY 80  
112 NGVSIADPINTNKWAAVAVGMTNSWDPNDGAILRSSDQATWQITPLP---FKLGNN 168  
81 MGVLVAALBPSDDKRIYAMTKYTTQW-AGYGAIIILSEDDGETWTVNLQKYGIGNE 139  
169 PGRGGERLAVDNNNDIILYFGAPSGKGLMRSTDGATWSQMTNPDVGTIYANPTDTTG 228  
140 DGRNAGERLQVDENPFSVLPMGT-TKYGWKSEDFGNWKVKVDSFST----- 186  
229 YQSDIGYVWVAPDKSSSLGQASKTIIFGVADPNNPVFWSRDGGATWQAVPGAPTGTIP 288  
187 -----SVTFVLEDEKSGEKSPTRPIFVGCSAPKG-IFVTEDEGTTWNLPLPNDLIP 239  
289 HGVFPDVNHVLYATNSNGPYDSSGDVWKFVSTSGWTRISPVSTDTANDYFGYSG 348  
240 LRKQIH-DEILVYTSNALGPNGATRGAVMKYVIAQKKYDVTYPMKD-----FGYCG 291  
349 LTDRQHPNTIMVATQISWMPDTIIFRSTDDGATWTRIMDTSYPNLSRLVLDISAEPN 408  
292 IDVQ---HWIVYSTIDRWYPHDEIFLSLNGGFTWPLEKANP-----DINKAP 339  
409 LTFGVQPNPPVPSPKLGMDDEMAIDPNSDRMLYGTGATLYATNDLTKMDSGGQIHAP 468  
340 IK---DLNP-----HWISD-VKIDPDMNRAIFTTGXYGWTVEIKKSFEG---MGR 384  
469 MYK-----GLEETAVNDLISPPSGAPLISALDGLGFTADVTAVPSTIFTSPVFTGT 522  
385 PVKMIENRGLLEETVVLQVLPPIGERPPLSALADWGGFRIESIDTPSSMY-PLKMTSL 443  
523 SDVYAEALNPSIIVRAGSPDSSQPNDRHVAFTDGGKNMFO-GSEPGCVTGG--TVAA 579  
444 GIAFAYQNSKFVARYTY---TYP---FLSYSDGGINMEIETVBEIGTIDGRLSIAVS 497  
580 ADGSRFVMAPGDGPVYVAVFGNSMAASQGVPA---NAQIRSDRVNPTFYAL--SNG 634  
498 NDGKTLVMSBAN--HEVIVSDDGKSKKALISVPVEFNYPFASDPVNPSEKFYIFPMKNG 555  
635 TFRSTDDGVTTPQVAAGLPSSGAVGMFHA---VPGKGGDMLAAS-SGLYHSTNGS 689  
556 DFLISDXGSEFKK-GAKLPSFDNMWVSLSYFVLVLPDRREGDITMLAQNMGLRSKDG 614  
690 SMSAITGVSSAVNVGFGKAPGSSSYPAVFVGTIGTVGAYRSDCGTMTVLINDPOHY 749  
615 TFERLGAVDIAIYVIGAPRPGTDPALYIANGVNVGYTIGFMSTDEGKTMRIINDKQGF 674  
750 GNVGQAITGDHANLARVYIGTNGRGIYVADI 780  
675 G-WIHIMIDMNEFRGILPLGTBERGIIIVEV 704

RESULT 5  
T31085  
xylanase - Caldicellulosiruptor sp.  
C/Species: Caldicellulosiruptor sp.  
C/Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
C/Accession: T31085  
R/Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.  
submitted to the EMBL Data Library, December 1997  
A/Description: Family 10 and 11 xylanase genes from Caldicellulosiruptor sp. Rte69B.1.

A:Reference number: Z20972  
 A:Accession: T31085  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1779 <MOR>  
 A:Cross-references: EMBL:AF036924; NID:g2760905; PID:g2760909; PIDN:AAB95326.1  
 C:Genetics:  
 A:Note: xymc

Query Match 8.4%; Score 431.5; DB 2; Length 1779;

Best Local Similarity 19.8%; Pred. No. 5.7e-12; Mismatches 382; Indels 477; Gaps 53;

Matches 252; Conservative 162; Mismatches 382; Indels 477; Gaps 53;

65 VDCIVFEGAGPGLIYV-----RTDIGMYRMDANGRMWIFPLDWMGNMNGVGVSIAD 120  
 20 VGLTIHQEAKAAAYTVDFEGDTLSFFAYGKSNTA-----VDMGN-AYNGKSIRRS 71  
 121 PINTKVAAYAGMTYNSWMDPDGAILRSSDQATWQIT-----PLPFLK-----G 165  
 72 -NRSSIMDGVAV-----DVKNIMNNGTWWVSAVYGHYSQKPAFGISAAYDDG 119  
 166 GNMGRGGEGLANDPNNNDILYFGAPSGKGLMRSTDSGATWSQMTNPFVGTIYANPTD 225  
 120 SGVKSSTLIGEVVAL-PN-----YWKJL-----VGRKMTPNISN 150  
 226 TTGYQSDIDGVVWVAFDKSSSLGQASKTIIFVGVADPN---NPFWSR--DGGAT--WQ 277  
 151 VRNLLIYVHTIVESVDYNVD-----YQIMDDNSYLSNATYFSSGFSSTTEGQ 201  
 278 A-----VPGAPTFGLPHKGVDPVNHVLYA--TSNTG--PYDG--SSGDVWKFVSV-- 323  
 202 ARGSGVTVKPDVSVVANGKYS-----LVYSGRTSMWHGQIIPVDTLLEGQKYSKLSVWY 256  
 324 -TSGTWIRISVPSSTDYAND-YFGISGLTDQHPNTINVAQISWMPDTIIFRSIDGA 381  
 257 QNSGSTOKMSLTMRFRATIDPSTYENILYNDVDSNTWVBSGSY-----SIPAGV 308  
 382 TWITIMDTSYFNSRLRY-----VLDIS--AEP----- 407  
 309 TVSELILLYEQAQNNALFAWDDDKIYDLSKLAPEMEIISLEKXKDYKGVVALSYKSI 368  
 408 -----WLTGVOQNPVPSPKLGW 426  
 369 ASDTEKKMVLKHPNSITAGNEMKPELLISENNYNSKADDEVNFATSNINIAIRHTLW 428  
 427 MDEA-----NAIRP----- 435  
 429 HEQTPDWEFFKDANGTLKSKALLSLKQYITVVGKGVYAMDVVNEAIDESQNGFR 488  
 436 ----FNSDRMLYGTGATLYA-----TNDLTRKMDSGQIHTAPWKGLBETAVNDL 481  
 489 RSNWYNICGPEYIEKAFITWHEADPAKFLYNDVYNTENQKQOFIYNMKSLE----- 542  
 482 ISPSGAPLISALGDLGGFTTHADVTAVPSTIFTSVFTT--GTSVDYAEALNPSIIVRAGS 539  
 543 ----KGVF-INGIG-LQCHINDWPSISIENTIKLFSIGLEIHTELDMSPFGWGS 596  
 540 FDPSSQPD-----RHVAF-----STDGKKN--WQ 563  
 597 TSVSTPRDLLIKQAMRYKELFDLFKKNVITNVTFWGLKDDYSWLSQMFGRSDYPLD 656  
 564 GS-----EEGCVTTGGTV-----AASA 580  
 657 GNYSKAFWMLLEPTVVPVNSTLPAAPAIQVPTPTSTPTPTPTVSAITPPAPTASPA 716  
 581 DGSRFVWAPDPCQPVVAVGFGNSMAASQGVAPANAQIRS-----DRVAP 625  
 717 GGS--YWPFSF-----SYGALKTWYANGMNSTTNVNLPRKIKIENVGSTAVDLSRVY 767  
 626 KTFYALNSGTYYRSTDDGVTFOP-----VAAGLPSSGA-----VGWHEHAVP 667  
 768 RYWTYI-DGEAAQSVSVASSINPAYIDVYVVLGANAAGADYVEVGFGRSGAVILAAGS 826

QY 668 GKEGDLMLAASGLYHSTNGSSWSA-----ITGVSSAVVYGEKSAFGSSYPANFV 719  
 DB 827 YKEIRLSICKSSGSYNQSDNYSVRANSYIENEKYGYIDDVLYWGERPERNAQIKWYA 886  
 QY 720 VGTIGGVTA-----YNSDCCGT-----WYLINDHQYQWGAQIGDHANLR 764  
 DB 887 NGMLSSFTVNLNPKIKIENGTAVDLSRVKRYWTTIDGATQSVSVASSINPAYIDVR 946  
 QY 765 RYVIGTNGRGIYV-----GDIGGA-----PSSG----- 787  
 DB 947 VYKLGANAGADYVVEVGFGRSGAVILAAGSTKEIRLSICKSSGSYNQSDNYSVRANSY 1006  
 QY 788 -----PSPVSPASPSLSPPSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 828  
 DB 1007 IENEKYTYIDVLYWKEPSSRGTKPAG--EYTPAPPTPTPTPTPTPTPTPTPTPTPT 1064  
 QY 829 SP 883  
 DB 1065 TVTATPT 1124  
 QY 884 PGDNQIKPGLOVYNTSSSVYDLSYTVYKWFTRDGSSTLVYNCDMAAIGCNIRASFGS 943  
 DB 1125 ASYGSIRPWFKTYNGSSSVYDLSRVKIRYWTVDGDKPGSAV-CDMAQIGASVNTFRVX 1183  
 QY 944 VNPATPTADTYIQ 956  
 DB 1184 LTRGVSGADYYLE 1196

# RESULT 6

T31337 1,4-beta-glucanase (EC 3.2.1.-) - Anaerocellum thermophilum (fragment)

C:Species: Anaerocellum thermophilum

C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000

C:Accession: T31337

R:Zverlov, V.; Mahr, S.; Riedel, K.; Bronnenmeier, K.

Microbiology 144, 457-465, 1998

A:Title: Properties and gene structure of a bifunctional cellulolytic enzyme (Cela) from

omains.

A:Reference number: Z21003; MID:96154434; PMID:9493383

A:Accession: T31337

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1711 <ZVE>

A:Cross-references: EMBL:Z86105; NID:e1071329; PID:e350354; PIDN:CAB06786.1

C:Genetics:

A:Gene: cela

C:Keywords: glycosidase; hydrolase

Query Match 7.8%; Score 398.5; DB 2; Length 1711;

Best Local Similarity 22.9%; Pred. No. 1.7e-10; Indels 281; Gaps 40;

Matches 197; Conservative 116; Mismatches 267; Indels 281; Gaps 40;

262 PNNPVWMSRDGATWQA-----VPGAPTFGLPHKGVDPVNHVLYIATSNTGPGYGS--S 315  
 25 PN-----WVNN-----NMRGDSALKQDQNGLDLTGWMFADGHWKF-----NLPMSTGTMLS 73  
 316 GDYWKFS--YISGTRISVPSTDTANDYFGYSGLTIDRQHPNTIMATQI----- 365  
 74 WAYVEYKDAFVKSGOLEHI--LNOIEWVNDYF-----VKCHPSKVVYVYQVVDGSKDH 124  
 366 SWW-----PDITIFSTDGAMTWTRIMWTSYPNKSLRVLDLSAEPMLTFGVOP 415  
 125 AMWGPAEWMQMERPSFKVYDSSPGSTVTE--TAASLAASIVLK-----DR 169  
 416 NPVPSPKLGWMDMAAIDPFNSDRMLYGTATLYATNDLTAKMDSGQIHTAPWKGLB 475  
 170 NPTKATYLOHAK-----LYBAEYTKBDAG-----Y 197  
 476 TAVNDLISPSG-----APLISALGDLGGFTTHADVTAVPSTIFTSVFTTGTSDVY 526  
 198 TAANGYVNSWSPGYDELMAAVMLYLATNDSTYLTKEA-----SYQNWPKISGSNTIDY 252

QY 527 -----AELNPSIIYRAGSPDSSQPNDRHVAFTDG-----GKNMFG 563  
DB 253 KWAHCMDVHNGAAL---LAKITGKDIYKQIIESHLDYWTGNGERIKYTPKGLAMLD 309  
QY 564 --GSEPGVTTGTVTAASAD-----GSRFWAAPGDPGPVYAVGFG-- 603  
DB 310 QMGLRATYTTAFLAFVYSIDWVGCPSTKKEIYRKGESEQIDYALGSAGNS--FVVGFGTN 367  
QY 604 -----NSWAASQGVPA-----NAQIRSD-- 621  
DB 368 PPKRPHRTAHSWADSQSIPSYHHTLYGALVGGPGSDSYTDDISYNNVACADYNA 427  
QY 622 -----RVNP---KTFFALSNCTFYRSTDGCVTPQVPAAGLPSGAGVGMFH 664  
DB 428 GFVGALAMVQLYGNGNLPDFKALERTNDEFF-----VEAGINAGTNPFEIK 476  
QY 665 AVPKBEGDMLAAS-----GLYHSTNGSSWSAITGVSAVNVFGKAP----- 710  
DB 477 AIVNNQSG-WPAKADKLFKRYFTVDLSLIRAGYSPNQLT-LSTYNNQAKVSGPYVMDA 534  
QY 711 --GSSYPAVFVVGIT--GVTGAVRSDCGTTWVLINDOHQYGMWGAITGDHANLRV 766  
DB 535 SKNIYIIVNDPTGTLIYVGQDKYKE-----VQFRIAPQNVGMNDNSFQDIKGV 588  
QY 767 YIGT-----NGRGIYVDIGAPSGSPSPSVSPASPSLSPSPSSPSFSP 816  
DB 589 SSGSVVTKTYIPLYGDVKKVWGDGPTSGATPTPTA-----TATPTPTPTPTPTPT 642  
QY 817 SSSSSSSPSPPSP 875  
DB 643 TS--TATPTPTPTPTPTPTPTPTPTPTA--TPTSTPTPTST-----TPTSTVAGQIV 690  
QY 876 QYKNNDAFPGNQIKPGQOVNVTGSSVDLSTVTVRWFTFRDGSSTLYVNCDMAIIGCG 935  
DB 691 LYAKETNSTNTIRPMLKVNTGSSSIDLSKVTIRYVTVVGDRAQSAIS-DMAQIGAS 749  
QY 936 NIRASFGSVNATPTADTYLQ 956  
DB 750 NVTPEFVTLSSSVSGADYYLE 770

## RESULT 7

117120  
cellulase (EC 3.2.1.1) precursor, thermoactive - Caldocellum saccharolyticum  
C:Species: Caldocellum saccharolyticum  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-Nov-2000  
C:Accession: 117120; A43745  
R:Te'o, V.S.; Saul, D.J.; Bergquist, P.L.  
Appl. Microbiol. Biotechnol. 43, 291-296, 1995  
A:Title: CelA, another gene coding for a multidomain cellulase from the extreme thermoph  
A:Reference number: Z18658; MUID:55336703; PMID:7612247  
A:Accession: 117120  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1742 <EEO>  
A:Cross-references: EMBL:132742; NID:G537499; PID:G537500; PIDN:AAA91086.1  
R:Luethi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.  
Appl. Environ. Microbiol. 57, 694-700, 1991  
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding  
A:Reference number: A43745; MUID:91247819; PMID:2039230  
A:Accession: A43745  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1516-1544, 'A', 1546-1742 <LUE>  
A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72860.1; PID:g144293  
C:Genetics:  
A:Gene: celA  
C:Keywords: glycosidase; hydrolase

Query Match 7.7%; Score 397.5; DB 2; Length 1742;  
Best Local Similarity 21.7%; Pred. No. 1.9e-10;  
Matches 189; Conservative 109; Mismatches 269; Indels 303; Gaps 35;

QY 262 PNNPVMSRSDGATMQA-----VGAFTGFIPIHKGVFDPVNVHLV-LATNTGGPYDSSG 316  
DB 48 PN-----WVRN---NMGDSALKDQDNGDLTGCGWPDAGKFXKFLPMSYTGMLSWAAY 100  
QY 317 DVKFSVTSCTWTRIPVPSTDPANDYFGYSGLTIDROHPTNTMVAIOI-----SMV- 368  
DB 101 EYDQAVKSGQLEHT--LNDIEVNDYF-----VKCHSKVYVYQVGDGKDHAWMG 151  
QY 369 -----PDTIIFSTDGAATWTRIMWTSYPNLSRLRYLDISAPMLTFGVQPNPV 419  
DB 152 PAEWQMBERSFVYQSSPSGA-----VVAETAASLAASIVLKDRN 193  
QY 420 PSEKLGMDNBAALIDPNSDRMLYGTGATLYATNDLTKMNSGGQIHIAFWKILEEYAVN 479  
DB 194 PTAAYTLQHA-----KDIYFAEYVKSQSG-----YTAAN 224  
QY 480 DLSPSPG-----APLISALGDLGFTHADVNAVPSITFTSPVFTTGSVDY----- 526  
DB 225 GYTNSSGFTDELSMAVWLYLATNDSTYLTKE--SYQNPWKISGNSIIDIYKMAH 279  
QY 527 -----AELNPSIIYRAGSPDSSQPNDRHVAFTDG-----GKNM----- 561  
DB 280 CMDVHNGAAL---LAKITGKDIYKQIIESHLDYWTGNGERIKYTPKGLAMLDQWGS 336  
QY 562 -----FGSEPGVTTGTVTAASADG--SRFWAAPGDPGPVYAVGFG----- 603  
DB 337 LRYATTTAFLAFYYSWMSGCTGKKEIYRKGESEQIDYALGSTGRS--FVVGFGTNPFR 394  
QY 604 -----NSWAASQGVPA-----NAQIRSD----- 621  
DB 395 PHRTAHSWADSQSIPSYHHTLYGALVGGPGSDSYTDDISYNNVACADYNAFVG 454  
QY 622 -----RVNP---KTFFALSNCTFYRSTDGCVTPQVPAAGLPSGAGVGMFHAVPG 668  
DB 455 ALAKMVLKGNPIPFKALERTNDEFF-----VEAGINAGTNPFEIKAL-- 501  
QY 669 KEGDLWLAASGLYHSTNGSSWSA-----ITGVSAVNVFGKAPGSSYPAYF 718  
DB 502 -----VNNQSGMPARATNKLKFRYFVDLSLIRAGYS-----PNQL 537  
QY 719 VVGIT---GVTGAVRSDCGTTWVLIND-----DOHQYGMWGAITGDHANLR 765  
DB 538 TLSTNNQAKVSGPYVWDSNRNIYYIIVNDPTGTLIYVGQDKYKKEVQFRIAPQVQO- 596  
QY 766 VYIGTNGRGIYVDIGAPSGS-----PSPSVSPASPSLSPSPSPS 807  
DB 597 ---WDSNDSYFQDINGVSSGVTKTYIPLYEDIKWGESEBFTS-----GVSPYFAS 648  
QY 808 SPSPPSP 867  
DB 649 VTFPTPTPTPTA-----TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPVST 700  
QY 868 P-VSGGVQYKNNDAFPGNQIKPGQOVNVTGSSVDLSTVTVRWFTFRDGSSTLYVN 926  
DB 701 PATSGQIKVLYAKETNSTNTIRPMLKVNTGSSSIDLSRVIRYVTVVGDRAQSAIS 760  
QY 927 CDMAAIGCGNIRASFGSVNATPTADTYLQ 956  
DB 761 -DMAQIGASNVTFKFKVTLSSSVSGADYYLE 789

## RESULT 8

S50755  
hypothetical protein VSP-3 - Chlamydomonas reinhardtii  
C:Species: Chlamydomonas reinhardtii  
C>Date: 14-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 21-Jul-2000  
C:Accession: S50755  
R:Moensner, J.P.; Molendijk, A.J.; van Egmond, P.; Kils, F.M.; Goodenough, U.W.; Harting,  
Plant Mol. Biol. 26, 947-960, 1994  
A:Title: Domain conservation in several volvocalean cell wall proteins.  
A:Reference number: S50754; MUID:95093034; PMID:8000007  
A:Accession: S50755  
A:Status: preliminary



R; Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1991  
C;Genetics:





RESULT 15  
T34434  
hypothetical protein K06A9.1a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T34434  
R:Geisels, C.; Gattung, S.  
Submitted to the EMBL Data Library, December 1996  
A:Description: The sequence of C. elegans cosmid K06A9.  
A:Reference number: 221525  
A:Accession: T34434  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2232 <GB>  
A:Cross-references: EMBL:U80846; PIDN:AACT0890.1; GSPDB:GN00028; CESP:K06A9.1a  
A:Experimental source: strain Bristol N2; clone K06A9  
C:Genetics:  
A:Gene: CESP:K06A9.1a  
A:Map position: X  
A:Insertions: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1  
Query Match 5.7%; Score 290.5; DB 2; Length 2232;  
Best Local Similarity 21.8%; Pred. No. 1.6e-05;  
Matches 239; Conservative 114; Mismatches 428; Indels 315; Gaps 45;  
QY 18 SLAATASFAVAALGVLPPIA---ITASPAHAATQPYTMSNAI---GGGFGVDGIVFNE 72  
DB 235 SLGTSSPLPSISITSAIPASASSSPASSTTPVLLSSSTIQSSGTPPSSVASSP 294  
QY 73 GAGILYRTIGGMYRDANGRMIFLIDVGMNMGVINGVSIAD-----PINTMK 126  
DB 295 STVG-----STSG-----AASSSYATVSTIAGSTGTTTPVPGSS 330  
QY 127 VMAIVGWYTNMWDPNDAILR--SSDQATWQITP-LPEKLGAMP---GRGWERLAVD 180  
DB 331 --STTGSTPSSSSSGTSGTSTGTVTVVPGSSSTPSSSTPIASSSSPGSTVYA 388  
QY 181 PNNNNILYFGAPSGKGL---WRSTDGATWSQMTNFPDGVYIANPTDTGYQSDIQGVV 237  
DB 389 PGSSSTYGSTPSSASSSSSGTSGTSTGTVTVVPGSSSTPSSSTPIASSSSSGSTVT 448  
QY 238 WVAFDK-----SSSLGQAS-----KTIIVGVADPNNPVFWSRDGATWQA 278  
DB 449 VVSGSSSTYGSTPSSASSSAGTASTIGSTGSTATIVPGSS-----SSVGSSTQGA 500  
QY 279 VPGAPGTGIRPKGVDPNNHLYIATSNCGPYDSSGCVKFSVTSGTWTSISPSTPD 338  
DB 501 SPSSPG-----TMSVSGP-TGSTVTVVPGSSSTSPA-PSSSPNPS 540  
QY 339 TANDYFGYSGELTDROHNTIMVATQISWMPDTIIFRSTDGATWT--RIMDMTSYPMR 395  
DB 541 PAS-----TGSTTTTSSSSIIIVT-----VGSITVSGSTGTSOSTLASSTATPGS 586  
QY 396 SLRYVLDISAE-----PMLTFGVQPNP--PVSPKLGWMDZAMADDPNS 438  
DB 587 SSTVPSSSSPQSSSPAPNTGSTTPSQTSQSPSPMNPSSSTPTGSSQSTITTEPGSTA 646  
QY 439 DRMLYGTGATLYATNDLTKMDSGGQIHIAEMVKGLEETAANDLIPPSGAPLISALGDL- 497  
DB 647 SSPGSGTSGTSPVAITEVTSQST-----VPSGSSIGTQSTNPSPPSSISPSSTGMSLT 700  
QY 498 -----GGFTHADVTAVPS-----TIFSPVFTTCTSDVYALNPSIIIVRGS 539  
DB 701 SEPSPSSSTQSSGAGSTLTTPSPNPQSTSSLESSTSGATSSGAGTTWTSPPQSSSVGS 760  
QY 540 FDPSSQPNDRHVAFTDGAKNWFGSS--EPG-GVTTGGTVAASADGSRFVWAPG----- 590  
DB 761 SQGSTSP-----AASTSGEMTSQSTQTPGSSVSTSAAILSTQOOSVSTNPGSTVTRP 815  
QY 591 -----DGGQPVVY---AVFGNSMAASQGVANAOIRSDRVNPKTFYALSNGTFRS 639  
DB 816 STVGSSTSSGSTVTVGSTEASTGSSVASSSPAPSTSQ-----NPNPSTSSGSSMITQS 869

QY 640 TDGVTFFQPVAAGLPSSGAVGVMFHAVPGKEGDLMLAASGLYHSTNGSSNSA-----IT 695  
DB 870 PYPSSQSTSPVESSTPS-----PGSPGTLTSTSPSPSQSTTIGSTGSTSPGIS 919  
QY 696 GYSSAANVFGKAPGSSYPAVFVVGITGVTAAYRSDDCGTTWVLINDQHQYGMWQA 755  
DB 920 TTSEMTSQSTQTPPSTGTVTQPSVSDT----- 951  
QY 756 ITGDHANLRVYIGTNGRGIVYDYGAPSGSPSPVSPASPSLSPSPS-----PSSSP 810  
DB 952 -----SSGSTVTVGSTEQ--SSPISPTSGVNTNPTSSGSSMSTQTPQSSQ 995  
QY 811 SPSP--SPSSSPSSSPSP-----SPSPSPSP-----SNP-----SPASSPSS 848  
DB 996 STSPVSSSTGATSSSGSPGTTLTSLSPSPSPSTIGSQGSTSPVVTISOGSTETPGS 1055  
QY 849 SPSPSSSP-----SSSPPTSSSPVSGVAVQYKND 881  
DB 1056 TGSTVTKPSTVSGSASAGSTATMGSTEASTSGSSSTSPNPQSTSPSTGA-----T 1108  
QY 882 SAPGDNQIRKGLQVNVNTGSSVDLSTVTVVWFTPRDGSSTLYNMC--DMAIGCNIRA 939  
DB 1109 SSPGSS-----GTLTSLSPSPSQSSTIG-----SSQGSTSPVVTSTSGDWTSGSTQIPG 1159  
QY 940 SPGS--VNPATPTADT 953  
DB 1160 STGSTVTPSTGSGST 1175

Search completed: May 14, 2004, 09:18:40  
Job time : 32 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2004, 09:10:23 ; Search time 18 seconds  
(without alignments)

2768.397 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135  
Sequence: 1 MDRSENIRLTWRSRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	397.5	7.7	1742	1 GUNA_CALSA	P22534 caldocellium
2	340.5	6.6	1331	1 MAMB_CALSA	P22533 caldocellium
3	293	5.7	279	1 Y091_NPVOP	O10341 oryza pseu
4	290	5.6	1039	1 GUNB_CALSA	P10474 c endogluc
5	283.5	5.5	474	1 VTP3_TTVIV	P19275 thermoprote
6	272	5.3	555	1 GPl_CHIRE	Q94P66 chlamydomon
7	261.5	5.1	1853	1 CIPR_CLOTM	Q06851 clostridium
8	250	4.9	268	1 NO20_MEDTR	P93329 medicago tr
9	245	4.8	772	1 CIPB_CLOTM	O01866 clostridium
10	245	4.8	879	1 GUNI_CLOTM	O02834 clostridium
11	244	4.8	1664	1 SLB1_CLOTM	Q06852 clostridium
12	235.5	4.6	2124	1 PGCA_PAT	P07897 rattus norv
13	222.5	4.3	360	1 VTPX_TTVI	P19274 thermoprote
14	222.5	4.3	5703	1 MUSB_HUMAN	Q09484 homo sapien
15	221	4.3	2132	1 PGCA_MOUSE	P10220 herpes simp
16	218.5	4.3	3164	1 TBGV_HSV11	P10220 herpes simp
17	216	4.2	449	1 APG_BRANA	P40603 brassica na
18	214	4.2	532	1 SFG7_DICDI	P22688 dictyostell
19	210.5	4.1	545	1 GUN1_ACICR	P54583 acidotherru
20	209	4.1	535	1 SPRC_SYNY3	P74745 synochocyst
21	209	4.1	2415	1 PGCA_HUMAN	P16112 homo sapien
22	208.5	4.1	485	1 SSGP_VOLCA	P21997 volvox carc
23	207.5	4.0	544	1 GPlO_DICDI	Q06885 dictyostell
24	205.5	4.0	700	1 GUNA_PAEIA	P10475 paenibacill
25	204.5	4.0	499	1 GUN2_BACST	P10475 paenibacill
26	204.5	4.0	534	1 APG_ARATH	P40602 arabidopsis
27	204.5	4.0	1970	1 RPB1_HUMAN	P24928 homo sapien
28	204.5	4.0	1970	1 RPB1_MOUSE	P08775 mus musculu
29	204	4.0	444	1 GUNN_ERWCA	Q59393 erwina car
30	203.5	4.0	499	1 GUN1_BACST	P07983 bacillus su
31	201	3.9	456	1 GUNA_MICBI	P26414 microbispot
32	199.5	3.9	499	1 GUN3_BACST	P23549 bacillus su
33	199.5	3.9	1367	1 AMTH_YEAST	P08640 saccharomyc

34	199	3.9	1324	1 IRS2_HUMAN	Q94H2 homo sapien
35	198.5	3.9	505	1 GUNV_ERWCA	Q47036 erwina car
36	198	3.9	467	1 RPB1_CRIGR	P1414 cricetus
37	197	3.8	344	1 YBIE_SCHPO	P81719 schizosach
38	197	3.8	2333	1 PGCA_CANPA	O28343 canis fami
39	195.5	3.8	1419	1 ALAI_CANAL	O13366 candida alb
40	194.5	3.8	1852	1 RPB1_CABEL	P16356 caenorhadi
41	194	3.8	1150	1 APMT_PIG	P12021 sus scrofa
42	193.5	3.8	1260	1 ALSI_CANAL	P46590 candida alb
43	192.5	3.7	902	1 RPB1_DICDI	P35084 dictyosteli
44	192	3.7	2364	1 PGCA_BOVIN	P13608 bos taurus
45	191.5	3.7	412	1 CSP_FLARA	P02893 plasmodium

## ALIGNMENTS

RESULT 1  
GUNA\_CALSA STANDARD; PRT; 1742 AA.  
ID GUNA\_CALSA  
AC P22534;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)  
DE (Cellulase A).  
GN CELA.  
OS Caldocellum saccharolyticum (Caldicellulosigranulosa).  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;  
OC Caldicellulosigranulosa.  
OX NCBI\_TaxID=44001;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95336703; PubMed=7612247;  
RA Te O V.S., Saul D.J., Bergquist P.L.;  
RT "Cela", another gene coding for a multidomain cellulase from the  
RT extreme thermophile Caldocellum saccharolyticum";  
RL Appl. Microbiol. Biotechnol. 43:291-296(1995).  
RN [2]  
RP SEQUENCE OF 1516-1742 FROM N.A.  
RX MEDLINE=91247819; PubMed=2039230;  
RA Leuthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;  
RT "Cloning, sequence analysis, and expression in Escherichia coli of a  
RT gene coding for a beta-mannanase from the extremely thermophilic  
RT bacterium 'Caldocellum saccharolyticum'";  
RL Appl. Environ. Microbiol. 57:694-700(1991).  
CC -1- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN  
CC ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL  
CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE  
CC CELLULOSE.  
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -1- PFM: The linker region (also termed "hinge") may be a potential  
CC site for proteolysis.  
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY  
CC E (FAMILY 9 OF GLYCOSYL HYDROLASES).  
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY  
CC L (FAMILY 48 OF GLYCOSYL HYDROLASES).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: L32742; AAA91086.1; -;  
CC EMBL: M36063; AAA72860.1; -;  
CC EMBL: L01257; -; NOT ANNOTATED\_CDS.  
CC PIR: T17120; T17120.  
CC HSP: P26221; 1TF4.  
CC InterPro: IPR001956; CBD\_3.

DR InterPro; IPR008965; Cellul bind.  
 DR InterPro; IPR000556; Glyco\_hydro\_48.  
 DR InterPro; IPR001701; Glyco\_hydro\_9.  
 DR InterPro; IPR008928; Glyco\_trans\_6hp.  
 DR Pfam; PF00942; CEM\_3; 3.  
 DR Pfam; PF02011; Glyco\_hydro\_48; 1.  
 DR Pfam; PF00759; Glyco\_hydro\_9; 1.  
 DR PRINTS; PR00844; GLHYDRASE48.  
 DR ProDom; PD011947; CSD\_3; 2.  
 DR ProDom; PD011903; Glyco\_hydro\_48; 1.  
 DR PROSITE; PS00592; GLYCOSYL\_HYDROL\_F9\_1; 1.  
 DR PROSITE; PS00598; GLYCOSYL\_HYDROL\_F9\_2; 1.  
 DR Cellulose degradation; Hydrolase; Glycosidase; signal; Repeat.  
 KM SIGNAL 1 23  
 FT CHAIN 24 1742 ENDOLUCANASE A.  
 FT DOMAIN 24 642 CATALYTIC 1.  
 FT DOMAIN 643 700 LINKER ("HINGE") (PRO-THR BOX).  
 FT DOMAIN 701 857 CELLULOSE-BINDING (BY SIMILARITY).  
 FT DOMAIN 858 903 LINKER ("HINGE") (PRO-THR BOX).  
 FT DOMAIN 904 1060 CELLULOSE-BINDING (BY SIMILARITY).  
 FT DOMAIN 1061 1112 LINKER ("HINGE") (PRO-THR BOX).  
 FT DOMAIN 1113 1742 CATALYTIC 2.  
 FT ACT\_SITE 396 396 BY SIMILARITY.  
 FT ACT\_SITE 434 434 BY SIMILARITY.  
 FT ACT\_SITE 443 443 BY SIMILARITY.  
 FT CONFLICT 1545 1545 T -> A (IN REF. 2).  
 SQ SEQUENCE 1742 AA; 193696 MW; 3F0699A2123BED07 CRC64;

Query Match 7.7%; Score 397.5; DB 1; Length 1742;  
 Best Local Similarity 21.7%; Pred. No. 4.7e-10;  
 Matches 189; Conservative 109; Mismatches 269; Indels 303; Gaps 35;

QY 262 PNNPFFMRDGCATGQA-----VPGAPTFPIPHKGVDPNPNHLY-LATNTGSPYGGSSG 316  
 DB 48 FN---WVRN---NWRGSAKXGQDNGDLTGCGWFDAGDHKEFNPMSYTGMLSWAA 100  
 QY 317 DVWKEFSVSGTWRTISPVSTPDANDYFGYSGLTIDRHPNTIMVATQI-----SW- 368  
 DB 101 EYKDAFVSGGLEHNI--LNQIEWANDYF-----VKCHPSKVVYVYQGDGDKDAWNG 151  
 QY 369 -----PDTIIFRSTDGATWTRIMWTSVPNRSIATVYLDISAEPMILTEGVQNPV 419  
 DB 152 PAEVQMERSPKVTQSSPSGA-----VVAETASIAAASIVLKDN 193  
 QY 420 PSPKLGMDENAIIDPNSDRMLYCGATLYATNDLTCKDSSGGHIAWVGLSETAYN 479  
 DB 194 PTKATYLOHA-----KDLYERAEVTKSDSG-----TYAAN 224  
 QY 480 DLISPPSG-----APLISALGDLGGFTADVTAVPSTIPTSPTFTGTSVDY---- 526  
 DB 225 GYVNSWGFYDELMAAVALYLATNDSTYLTAE-----SYQONWPKISGNSIIDYKMAH 279  
 QY 527 -----AELNPSITVRAGSFDPSSQPNDRHVAFTDG-----GKNW----- 561  
 DB 280 CMDVDVHNGAAL---LLAKITDKDYKQIIESHLDYTTGVNGERIKYTPKGLAMLQWGS 336  
 QY 562 -----FQSGRPGVTTGGTVAASADG-SRNVMAFGPGGQGVVAVAGG----- 603  
 DB 337 LRYATTTAFLAVYSPWSCPTGKKEVTKRFESQIDVALGSTRS--FVVGGTNPCKR 394  
 QY 604 -----NSWASQGVPA-----NAQIRSD----- 621  
 DB 395 PHHRTAHSSWASQSIPTSHRHTLYGALVGGPSSDSTYDDISNYNNVEACYNAGFVG 454  
 QY 622 -----RVNP-----KTFVALSNGTFYRSTDGVTQPVAAGLPSSGAVGVFAVPG 668  
 DB 455 ALAKMYLYGKPIPDFALETPTNDEFF-----VEAGINASGNFIETIKAI-- 501  
 QY 669 KEGDMLAASSGGLYHSTNGSSWSA-----ITGVSAVAVNGFGASAPSSSTPAVF 718  
 DB 502 -----VNNQSGWPARATNKLKRFYFVDSLILIKAGYS-----FNQL 537  
 QY 719 VVGTI---GGVTGAVRSDCGTTWVLIND-----DQHGYGMGQAITGDHANLR 765

DB 538 TISTNNGAKVSGPYWSSRNIIYILVDFGTLLIPGQDRIKKEVQPRIAPQVQ- 536  
 QY 766 VYGTNGRGIVYDIDGAPSS-----PSPSVSPASPSLSPSPSPS 807  
 DB 597 ---WDSNDSYSPDIDKVGSSGVVTKYIPLVDEDIKWEBEFGTS-----GVSPPTAS 648  
 QY 808 SSPSPSPSPSSSPSS 867  
 DB 649 VPTPTPTPTA-----TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 700  
 QY 868 P-VSGGVKQYKKNDAFGNQLKPGLOVNTSSSVLSTYVVRVFTFDGSSSTLVN 926  
 DB 701 PATSGQIKVLYANKETSTNTITRPMLVKVNSSSSSIDLSRYTIRVYVYVGERAGSAIS 760

QY 927 CDWAAIGCGNIRASPGSVNPAFTPTADTYQ 956  
 DB 761 -DWAQIGASVTRKFKVLSSVSGADYILE 789

RESULT 2  
 NAME CALSA  
 ID NAME CALSA STANDARD; PRT; 1331 AA.  
 AC P22533;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-JUL-1993 (Rel. 26, last sequence update)  
 DT 28-FEB-2003 (Rel. 41, last annotation update)  
 DE beta-mannosidase/endoglucanase A precursor [includes: Mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) (beta-mannanase) (Endo-1,4-mannanase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase)].  
 GN MANA.  
 OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;  
 OC Caldicellulosiruptor.  
 OK NCBI\_TaxId=44001;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93119139; Pubmed=1476429;  
 RA Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;  
 RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a multidomain enzyme.";  
 RL Appl. Environ. Microbiol. 58:3864-3867 (1992).  
 RN [2]  
 RP SEQUENCE OF 1-346 FROM N.A.  
 RX MEDLINE=91247819; Pubmed=2039230.  
 RA Luthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;  
 RT "Cloning, sequence analysis, and expression in Escherichia coli of a gene coding for a beta-mannanase from the extremely thermophilic bacterium 'Caldocellum saccharolyticum'.";  
 RL Appl. Environ. Microbiol. 57:694-700 (1991).  
 CC -FUNCTION: DEGRADATION OF HEMICELLULOSES, THE SECOND MOST ABUNDANT POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH MANNANASE AND ENDOGLUCANASE ACTIVITIES.  
 CC CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans, and galactoglucomannans.  
 CC -CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -MISCELLANEOUS: This enzyme is most active at pH 6 and 80 degrees Celsius.  
 CC -SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).  
 CC -SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL HYDROLASES).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announc/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; L01257; AAA71887.1; -  
 DR EMBL; M36063; AAA72861.1; -  
 DR PIR; A48954; A48954.  
 DR HSSP; Q06851; INBC.  
 DR InterPro; IPR001956; CBD 3.  
 DR InterPro; IPR008965; Cellul bind.  
 DR InterPro; IPR01547; Glyco\_hydro\_5.  
 DR Pfam; PF00942; CBM\_3; 2.  
 DR ProDom; PD001947; CBD 3; 2.  
 DR ProSite; PS00659; GLYCOSYL HYDROL F5; 1.  
 KM HydroLase; Glycosidase; Cellulose degradation; Signal;  
 KM Multifunctional enzyme.  
 FT SIGNAL 1 41  
 FT CHAIN 42 1331  
 FT DOMAIN 325 325  
 FT DOMAIN 326 361  
 FT DOMAIN 362 518  
 FT DOMAIN 519 564  
 FT DOMAIN 565 720  
 FT DOMAIN 721 1331  
 FT ACT\_SITE 162 162  
 FT ACT\_SITE 257 257  
 FT CONFLICT 338 338  
 FT CONFLICT 340 346  
 FT SEQUENCE 1331 AA; 146892 MW; 146892 MW; FFBGCA51BB8DF0E0 CRC64;  
 Query Match 6.6%; Score 340.5; DB 1; Length 1331;  
 Best Local Similarity 22.7%; Pred. No. 9.6e-08;  
 Matches 140; Conservative 97; Mismatches 189; Indels 191; Gaps 29;  
 426 WMDAMAIIDPNSDMLVTGATVATYNDLTGMDSGGQIHIAPMVYGLBETVNDLISPP 485  
 141 WKELKSVLDG-NEDFVITINIGNEPYGNNTYQW-----VNDTKN-- 178  
 486 SGADPLISAGDLGFTADVTAAPS--TFTSPVFTTGSVDYAEINPILVRA--GSF 540  
 179 ----AIKALRD-AGKHTITWDAPFMWGDMSTMTMDNAQINEMADPLRLVPSIMHYGV 233  
 541 DPSSQPNDRHVAFTSDG-----GKNWFGQSEPGVTTGGTVAASDGRFWAPEGDPGP 595  
 234 NTASKVEEYIKSFVDKGLPLVIGEFGHQHTD-----GDPDER 270  
 596 VV-----YANG-FCNSMAASGCVANNAQIRS--DRVNKTFFYALNNGTFYSTGG-- 643  
 271 AIVRYAQYKIGLFSWMSWGNSSVYGYLDVNMNDPNNPTPW-----GQWYKTNALIGTSS 325  
 644 -----VTFQPVAAGLP-----SSGAVGMFHAHPGKCDL--- 673  
 326 TPTPTSTVTPPT 382  
 674 --W-----AASSGL-----YHSTNGSSWSAITGV--SAVNGF-----GKSAP 710  
 383 IRPWLKVNSSGSSSIDLSRVTIRYWTVDGERAQSAISDMAQIGASNVTFKFLSSSVS 442  
 711 GSSYPAVAVVCTIGVTAYSDDCGTAVLINDP---QHQGN---WQALITGDHANLR 764  
 443 GADY--YLEIGFKSAGQLQPKDGEIQMRNKNKDMSNYNGNMSWISQMTS----- 494  
 765 RVTYIGTNGRGIVYGD---IGCAPSGSPSPSVSPASPSLSPSPSSPSPSPSPSPSPSPSP 820  
 495 ---YGENEKVATYIDGVLVWGOBPSGA-----TPAPAP 524  
 821 SSSP 879  
 525 TATPTPTPTVTPPT 576  
 880 NDSAPGDNQIRPGLOVAVNTGSSVDLSTVTVYVYVTRDGSSTLYNMCMAAIGCANIFA 939  
 577 KEIYSTVITIRIPWLKVNSSGSSSIDLSRVTIRYWTVDGERAQSAIS--DMAQIGASNVTF 635  
 940 SFGSVNPTPTPTADTYLQ 956

DB 636 KEVWSSSVSGADYTL 652  
 RESULT 3  
 ID Y091 NPVOP STANDARD; PRT; 279 AA.  
 AC 010341;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Hypothetical 29.3 kDa protein (ORF92)  
 OS Orygia pseudotsugata multicapsid polyhedrosis virus (OpMV)  
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
 OC Nucleopolyhedrovirus.  
 OX NCBI\_TaxId=164623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=7271300; PubMed=9126251;  
 RA Andrews C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,  
 RA Rohmann G.F.;  
 RT "The sequence of the Orygia pseudotsugata multinucleocapsid nuclear  
 RT polyhedrosis virus genome."  
 RL Virology 229:381-398(1997).  
 CC -1- SIMILARITY: TO CORRESPONDING ORF IN ACMPV.  
 CC -1- SIMILARITY: Contains 1 chitin-binding type-2 domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL; U75930; AAC59091.1; -  
 DR PROSITE; PS50940; CHIT\_BIND\_11; 1.  
 KM Hypothetical protein.  
 FT DOMAIN 213 273  
 FT SEQUENCE 279 AA; 29289 MW; 6FAADAA01009DBF0 CRC64;  
 Query Match 5.7%; Score 293; DB 1; Length 279;  
 Best Local Similarity 55.8%; Pred. No. 2e-06;  
 Matches 48; Conservative 26; Mismatches 12; Indels 0; Gaps 0;  
 783 APSGSPSPSVSPASGSLSPSPSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSA 842  
 DB 78 SPT 137  
 843 SPSPSSP 868  
 DB 138 SPT 163  
 RESULT 4  
 ID GUNB CALSA STANDARD; PRT; 1039 AA.  
 AC P10474;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Endoglucanase/exoglucanase B precursor [includes: Endoglucanase  
 DE (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)  
 DE (Cellulobiohydroxylase); Exoglucanase (EC 3.2.1.1) (Exocellulobiohydroxylase)]  
 DE (1,4-beta-cellulobiohydroxylase)].  
 GN CELB.  
 OS Caldicellum saccharolyticum (Caldicellulosiruptor saccharolyticus).  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;  
 OC Caldicellulosiruptor.  
 OX NCBI\_TaxId=44001;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89098398; PubMed=2789517;



[illegible]

AC 006851; DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cellulosomal scaffolding protein CfpA and a precursor (Cellulosomal  
glycoprotein S1/SL) (Cellulose integrating protein A) (Cohesin).  
GN CfpA.  
OS Clostridium thermocellum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1515;  
RN [1]  
RP SEQUENCE OF N.A., AND SEQUENCE OF 29-40 AND 42-43.  
RC SPRAIN-ATCC 27405 / DSM 1237;  
RX MEDLINE=93302508; PubMed=8316083;  
RA Demain A.L., Romaniec M.P.M., Kobayashi T., Hunkisson N.S.,  
RT "Sequencing of a Clostridium thermocellum gene (cfpA) encoding the  
RT cellulosomal SL-protein reveals an unusual degree of internal  
RT homology.";  
RL Mol. Microbiol. 8:325-334(1993).  
RN [2]  
RP SEQUENCE OF 1820-1853 FROM N.A.  
RX MEDLINE=93209931; PubMed=8458632;  
RA Fujino T., Beguin P., Aubert J.-P.;  
RT "Organization of a Clostridium thermocellum gene cluster encoding the  
RT cellulosomal scaffolding protein CfpA and a protein possibly involved  
RT in attachment of the cellulosome to the cell surface.";  
RL J. Bacteriol. 175:1891-1899(1993).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (1.25 ANGSTROMS) OF 184-321.  
RX MEDLINE=972838934; PubMed=90831077;  
RA Shimon L.J.W., Bayer E.A., Morag E., Lamed R., Yaron S., Shoham Y.,  
RN Frolow F.;  
RT "A cohesin domain from Clostridium thermocellum: the crystal  
RT structure provides new insights into cellulosome assembly.";  
RL Structure 5:381-390(1997).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.  
RX MEDLINE=97076134; PubMed=8914851;  
RA Tormo J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,  
RN Seitz T.A.;  
RT "Crystal structure of a bacterial family-III cellulose-binding  
RT domain: a general mechanism for attachment to cellulose.";  
RL EMBO J. 15:5739-5751(1996).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.  
RX MEDLINE=98022914; PubMed=9402065;  
RA Tavares G.A., Beguin P., Alzari P.M.;  
RN "The crystal structure of a type I cohesin domain at 1.7-A  
RT resolution.";  
RL J. Mol. Biol. 273:701-713(1997).  
CC -1- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSOME. IT  
CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE  
CC CELLULOLYTIC ENZYMES.  
CC -1- SUBCELLULAR LOCATION: Cell surface.  
CC -1- DOMAIN: THE COHESIN DOMAINS BIND TO THE DOCKERIN DOMAIN BORN BY  
CC THE CATALYTIC COMPONENTS OF THE CELLULOSOME.  
CC -1- SIMILARITY: Contains 9 cohesin domains.  
CC -1- SIMILARITY: Contains 2 dockerin domains.  
CC -----  
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L08665; -; NOT ANNOTATED\_CDS.  
DR EMBL; X67506; CAA47840.1; -;  
DR PIR; S36859; S36859.  
DR PDB; 1ANU; 23-JUL-97.



PDB; 1AOH; 08-JUL-98.  
 DR PDB; INBC; 26-SEP-97.  
 DR InterPro; IPR001956; CBD 3.  
 DR InterPro; IPR008965; Cellul bind.  
 DR InterPro; IPR002102; Cohesin.  
 DR InterPro; IPR002105; Dockerin\_1.  
 DR InterPro; IPR002048; EF-hand.  
 DR Pfam; PF00942; CBM\_3; 1.  
 DR Pfam; PF00963; Cohesin; 9.  
 DR Pfam; PF00404; Dockerin\_1; 2.  
 DR ProDom; PD001947; CBD 3; 1.  
 DR PROSITE; PS00018; EF\_HAND; UNKNOWN 1.  
 DR PROSITE; PS00448; CLOS\_CELLULOSE\_RPT; 2.  
 KW Cellulose degradation; Cell wall; Glycoprotein; Repeat; Signal;  
 3D-structure.  
 FT SIGNAL 1 28  
 FT CHAIN 29 1853  
 FT DOMAIN 29 182  
 FT DOMAIN 183 322  
 FT DOMAIN 323 363  
 FT DOMAIN 364 522  
 FT DOMAIN 523 559  
 FT DOMAIN 560 704  
 FT DOMAIN 724 866  
 FT DOMAIN 889 1031  
 FT DOMAIN 1054 1196  
 FT DOMAIN 1219 1361  
 FT DOMAIN 1384 1526  
 FT DOMAIN 1548 1690  
 FT DOMAIN 1791 1814  
 FT DOMAIN 1824 1847  
 FT CONFLICT 1615 1615  
 FT STRAND 185 188  
 FT STRAND 190 191  
 FT TURN 195 196  
 FT STRAND 198 206  
 FT TURN 210 211  
 FT STRAND 213 221  
 FT TURN 224 226  
 FT STRAND 227 234  
 FT TURN 236 237  
 FT TURN 243 246  
 FT STRAND 247 252  
 FT TURN 253 256  
 FT STRAND 257 263  
 FT TURN 265 266  
 FT TURN 270 271  
 FT STRAND 273 273  
 FT STRAND 277 286  
 FT STRAND 292 304  
 FT TURN 305 306  
 FT STRAND 309 309  
 FT STRAND 313 315  
 FT STRAND 317 319  
 FT STRAND 369 375  
 FT STRAND 381 382  
 FT STRAND 385 385  
 FT STRAND 387 393  
 FT STRAND 399 400  
 FT STRAND 401 403  
 FT HELIX 404 410  
 FT STRAND 418 428  
 FT TURN 430 431  
 FT STRAND 434 436  
 FT STRAND 438 440  
 FT HELIX 441 452  
 FT TURN 453 454  
 FT STRAND 455 463  
 FT STRAND 467 468  
 FT TURN 470 471  
 FT STRAND 473 482  
 FT TURN 483 484  
 FT STRAND 489 489

FT TURN 491 492  
 FT TURN 494 495  
 FT STRAND 498 498  
 FT STRAND 503 504  
 FT STRAND 509 512  
 FT TURN 513 514  
 FT STRAND 515 518  
 FT STRAND 1220 1224  
 FT STRAND 1226 1229  
 FT TURN 1231 1232  
 FT STRAND 1234 1242  
 FT TURN 1246 1247  
 FT STRAND 1249 1249  
 FT STRAND 1251 1257  
 FT TURN 1260 1262  
 FT STRAND 1263 1270  
 FT TURN 1272 1273  
 FT HELIX 1279 1282  
 FT STRAND 1283 1288  
 FT TURN 1289 1292  
 FT STRAND 1293 1299  
 FT TURN 1301 1302  
 FT TURN 1306 1307  
 FT STRAND 1309 1309  
 FT STRAND 1313 1322  
 FT TURN 1324 1325  
 FT STRAND 1329 1342  
 FT TURN 1344 1345  
 FT STRAND 1348 1348  
 FT STRAND 1351 1354  
 FT STRAND 1356 1360  
 SQ SEQUENCE 1853 AA; 196831 MW; 3ABDDC03ABFC5372 CRC64;

Query Match 5.1%; Score 261.5; DB 1; Length 1853;  
 Best Local Similarity 23.2%; Pred. No. 0.00033;  
 Matches 117; Conservative 74; Mismatches 159; Indels 155; Gaps 20;

QY 508 VPSTIFTSVPFTG---TSVDAELNPSIIVR-----AGSFDPSSQPNDRHYA--FSTDGG 558  
 DB 55 VPSKGMANCDPVLGYDPNVLVEVEVKGSIIKDPDPKSDSAIYPRKNIIVFLFAEDSG 114  
 QY 559 KMWFGSEPGVTTGGTVAASADGSRFWAPGDPGQPVYAVGFGNSMAASQGVPAQAOT 618  
 DB 115 R-----GTVAITQDG-----VPAT-----IVATV 133  
 QY 619 RSDRVNPKTFYALSNQFYRS-----TDGCV-----TFQPYAAGLPSGAY--- 659  
 DB 134 KSAAPAAITL--LEVGAFAQNDLVEISTTFVAGGVNIGSSVPTQP--NVPDGVVEI 188  
 QY 660 -----GVMFHAVPGKEGDLMLAASGCLYHSTNGGSSMSAITGVSSAVNYFG 706  
 DB 189 GKVTGSVGTVEIPVYFRGVPSK-----GIANCDFVFRYDPNVLVEIIGIDPDIIVDNP 244  
 QY 707 KSAAGSSYP---AVFVGTIGVGTGAYRSDGCTTWWLINDQHQYGNMGAITGD--H 760  
 DB 245 KSFDTAIYDPKTIIVFLFAEDSG--TGAY-----ATTKGVF 279  
 QY 761 ANLRVYIYGNKGIYVGDIGCAPSGSPSVSPASPSUSLSPSPSSSPSPSPSPSSP 820  
 DB 280 AKIRATVKSAPAGYIFDEVGFPAD-----NDLVQKXSFIDGYNVGNATPTKATP 332  
 QY 821 SSSP 880  
 DB 333 TMTATPTKSAIATPTR-----BSVPNTPTNTANTPVSGNLKVEFYNS 376  
 QY 881 DSAPGDNQIKRGQVNTGSSVDLSTVYRYWFTRDGSGSTLVYNCMAAI----- 932  
 DB 377 NPSDTNSINPQFVNTGSSAIDLSKTLRYVYTVVQDQKQTFW--CDHAALISGNSYN 435  
 QY 933 -GCNITASFGSVNPAITPTADTYIQ 956  
 DB 436 GITSNVKGFVWKSSSTNNADTYLE 460





DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer  
DB protein 1).  
GN OLPB.  
OS Clostridium thermocellum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1515;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCIB 10682;  
RX MEDLINE=93209931; PubMed=8459832;  
RA Fujino T., Begun P., Aubert J.-P.;  
RT "Organization of a Clostridium thermocellum gene cluster encoding the  
RT cellulosomal scaffolding protein Cipa and a protein possibly involved  
RT in attachment of the cellulosome to the cell surface.";  
RL J. Bacteriol. 175:1891-1899(1993).  
CC -!- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.  
CC -!- SUBCELLULAR LOCATION: Cell wall.  
CC -!- SIMILARITY: Contains 4 S-layer homology (SLH) domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X67506; CAA47841.1; -;  
DR PIR; T18262; T18262.  
DR InterPro; IPR008965; Cellul\_bind.  
DR InterPro; IPR001119; SLH.  
DR Pfam; PF00395; SLH\_3.  
DR PROSITE; PS01072; SLH\_DOMAIN; 2.  
DR Cell wall; S-layer; Signal; Repeat.  
KW SIGNAL 1 28 POTENTIAL.  
FT CHAIN 29 1664 CELL SURFACE GLYCOPROTEIN 1.  
FT DOMAIN 36 763 4 X 156 AA APPROXIMATE REPEATS.  
FT REPEAT 36 191 1.  
FT REPEAT 207 363 2.  
FT REPEAT 409 565 3.  
FT REPEAT 607 763 4.  
FT DOMAIN 771 1377 APPROXIMATE TANDEM REPEATS OF  
FT T-P-S-D-E-P.  
FT DOMAIN 1378 1449 GLY/PRO/SER/THR-RICH.  
FT DOMAIN 1453 1494 SLH 1 (INCOMPLETE).  
FT DOMAIN 1495 1565 SLH 2.  
FT DOMAIN 1566 1625 SLH 3.  
FT DOMAIN 1626 1646 SLH 4 (INCOMPLETE).  
SQ SEQUENCE 1664 AA; 178194 MW; 5F396695BA9FE74B CRC64;  
Query Match 4.8%; Score 244; DB 1; Length 1664;  
Best Local Similarity 20.8%; Pred. No. 0.0016;  
Matches 185; Conservative 96; Mismatches 316; Indels 292; Gaps 44;  
QY 100 LLDWVGNWNGYN---GVVSIADPTNTKVAAGVGMVYNSWDNDGAILRSSDQATW 155  
DB 174 LFDWDAETITGYEIQDPLVWEAEPLK-----ASVALELDKTKVKVGDITAT----- 223  
QY 156 QITPLPKLGNMGRGMLAVDPNDNI--LYFGAPSGKGLWRSTDSGATWSQMTNF 213  
DB 224 -----IKIENKMFAGYQLNICYDPTMLEAIELETGSAIAKRTWPT--GGTVLQSDNY 275  
QY 214 PDVGTIANPTDTGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPV-----FWS 269  
DB 276 -----GKTTAVANDV--GAGIINFAYSNLTKYRET--GVAEETGIIGKIGFRV 321  
QY 270 RDGAT-----WQAVPGATGPIPHKGVDPVNVHLVYATNSGTGPDYDSSGDWKFVS 323  
DB 322 LKAGSTAIRFEDTTAMPGAIEGTY---MFDWYG-----ENIKGYVSVQPGIEVAEGE 370  
QY 324 TSGTWTTRISVPSPSTDTANDYFGYSLGTIDRQHPTIMVATQI--SNWPDITIFRSTDCGAT 382

Db 371 EPGEPTTEPVP--TETPVD-----PTPTVTTEPVPSELDFSYVIMELD--KT 414  
QY 383 WTRIWDWTSYP-----NRSRLRYVLDISAEPWMLTFCGVQNPVPPSPKLGWDEAMAIDPF 436  
DB 415 KVKVGDIIITATIKIENKMFAGYQLNICYDPTMLEAIELETGSAIAKRTW-----PV 466  
QY 437 NSDRMLYGT--GATLYATNDLTKWDSGGQIHIAPMVKGL---BETA VNDLISPPSGAPLI 491  
DB 467 TGGTVLQSDNYGKTTAVANDV---GAGIINFAYSNLTKYRETGV-----AET 513  
QY 492 SALGDLG-----GFTHADVTAVPSTIFTSPVFT--TGTSV-DYAEINLPSIIVRAGSF 540  
DB 514 GIIGKIGFRLVKGASTAIRFEDTTAMPGAIEGTMYFDMYGENIKGYSVVQPGIEVAEGE- 572  
QY 541 DPSSQP-----NDRHVAFTSGGKNWFGQSGEGVGTGTGTVAA 578  
DB 573 EPTTEPVTETPVDPTPTVTTEPVPSELDFSYVIMELDKTK-----VKEGDVIA 622  
QY 579 SADGSRFVMAFGDPGQPVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALS----- 632  
DB 623 TIRVNNIKNLG-----YQIGI-----KYDPKVLFAFNIEGTDP 656  
QY 633 -----NGTFRSTD---GGVTFQPVAGLPSGAGVGMF--HAVPKGKGLMLAAS 678  
DB 657 IDEGTWPAVGGTILKNRDYLP TGVAINNVSIGILNFAAYVYFDDYREEGKSED----- 710  
QY 679 SGLYHSTNGSSWSAITGVSSAVNVGFG-----KSAPGSSYPVAFVVGTTG 724  
DB 711 -----TGIIG--NIGFRVLKADTTIRFELESMPGS----- 740  
QY 725 GVTGAYRSDDCGTTWVLINDQHOYGNWQATIGDHANLRRVYIGTNGRGIYVGDIGGAP 784  
DB 741 -IDGTYMLD-----WYL-----NRISG-----YVVIQAP 764  
QY 785 ----SGSPSPSVSPASPSLSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS 837  
DB 765 IKAASDEPIPTDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSD 824  
QY 838 PSPASPS 886  
DB 825 PTPSDEPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSD 872  
RESULT 12  
PGCA RAT STANDARD; PRT; 2124 AA.  
ID PGCA RAT  
AC P07897;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core  
DE protein) (CSPCP).  
GN AGC1 OR AGC.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88087070; PubMed=3693370;  
RA Doerge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;  
RT "Complete primary structure of the rat cartilage proteoglycan core  
RT protein deduced from cDNA clones.";  
RL J. Biol. Chem. 262:17757-17767(1987).  
RN [2]  
RP REVISION TO 698.  
RA Doerge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;  
RL J. Biol. Chem. 263:10040-10040(1988).  
RN [3]  
RP SEQUENCE OF 1856-2124 FROM N.A.  
RX MEDLINE=86250698; PubMed=2424893;  
RA Doerge K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;

Partial cDNA sequence encoding a globular domain at the C terminus of the rat cartilage proteoglycan.";  
 J. Biol. Chem. 261:8108-8111(1986).  
 -!- FUNCTION: This proteoglycan is a major component of extracellular matrix of cartilaginous tissues. A major function of this protein is to resist compression in cartilage. It binds avidly to hyaluronic acid via an amino-terminal globular region. May play a regulatory role in the matrix assembly of the cartilage.  
 -!- SUBUNIT: Interacts with FBLN1 (By similarity).  
 -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).  
 -!- DOMAIN: Two globular domains, G1 and G2, comprise the amino terminus of the proteoglycan, while another globular region, G3, makes up the COOH terminus. G1 contains link domains and thus consists of three disulfide-bonded loop structures designated as the A, B, B' motifs. G2 is similar to G1. The keratan sulfate (KS) and the chondroitin sulfate (CS) attachment domains lie between G2 and G3.  
 -!- PFM: Contains mostly chondroitin sulfate, but also keratan sulfate chains, N-linked and O-linked oligosaccharides.  
 -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 -!- SIMILARITY: Contains 4 link domains.  
 -!- SIMILARITY: Contains 1 C-type lectin family domain.  
 -!- SIMILARITY: Contains 1 Sushi (SCR) domain.  
 -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; M13518; AAA41836.1; --  
 EMBL; J03485; AAA21000.1; ALT\_SEQ.  
 PIR; A92623; A28452.  
 HSP; P98066; ITSG.  
 InterPro; IPR002353; AntifreezeII.  
 InterPro; IPR007110; Ig-like.  
 InterPro; IPR003006; Ig\_MHC.  
 InterPro; IPR003596; Ig\_V.  
 InterPro; IPR001304; Lectin\_C.  
 InterPro; IPR000538; Link.  
 InterPro; IPR003324; SCXXSG.  
 InterPro; IPR000436; Sushi\_SCR\_CCP.  
 Pfam; PF00047; Ig; 1.  
 Pfam; PF00059; lectin\_c; 1.  
 Pfam; PF02339; SGXXSG; 55.  
 Pfam; PF00084; sushi; 1.  
 Pfam; PF00193; Xlink; 4.  
 PRINTS; PR00356; ANTIFREEZEII.  
 PRINTS; PR01265; LINKMODULE.  
 ProDom; PD000918; Link; 4.  
 SMART; SM00032; CCP; 1.  
 SMART; SM00034; CLECT; 1.  
 SMART; SM00406; IGv; 1.  
 SMART; SM00445; LINK; 4.  
 PROSITE; PS00615; C\_TYPE\_LLECTIN\_1; 1.  
 PROSITE; PS00041; C\_TYPE\_LLECTIN\_2; 1.  
 PROSITE; PS50835; IG\_LIKE; 1.  
 PROSITE; PS00290; IG\_MHC; 1.  
 PROSITE; PS01241; LINK; 4.  
 Glycoprotein; Proteoglycan; Lectin; Signal; Sushi; Repeat;  
 Immunoglobulin domain.  
 SIGNAL 1 19  
 CHAIN 20 2124  
 DOMAIN 34 147  
 DOMAIN 170 247  
 DOMAIN 268 349  
 DOMAIN 504 581  
 DOMAIN 602 683  
 DOMAIN 1910 2036  
 POTENTIAL.  
 AGGREGAN CORE PROTEIN.  
 IG-LIKE V-TYPE.  
 LINK 1.  
 LINK 2.  
 LINK 3.  
 LINK 4.  
 C-TYPE LECTIN.

FT	DOMAIN	2040	2098	SUSHI.
FT	DOMAIN	48	140	G1-A.
FT	DOMAIN	152	247	G1-B.
FT	DOMAIN	253	349	G1-B'.
FT	DOMAIN	486	580	G2-B.
FT	DOMAIN	587	682	G2-B'.
FT	DOMAIN	685	798	KS.
FT	DOMAIN	801	1226	CS-1.
FT	DOMAIN	1227	1909	CS-2.
FT	DOMAIN	1910	2124	G3.
FT	DISULFID	51	133	BY SIMILARITY.
FT	DISULFID	175	246	BY SIMILARITY.
FT	DISULFID	199	220	BY SIMILARITY.
FT	DISULFID	273	348	BY SIMILARITY.
FT	DISULFID	297	318	BY SIMILARITY.
FT	DISULFID	509	580	BY SIMILARITY.
FT	DISULFID	533	554	BY SIMILARITY.
FT	DISULFID	607	682	BY SIMILARITY.
FT	DISULFID	631	652	BY SIMILARITY.
FT	DISULFID	1914	1925	BY SIMILARITY.
FT	DISULFID	1942	2034	BY SIMILARITY.
FT	DISULFID	2010	2026	BY SIMILARITY.
FT	DISULFID	2041	2084	BY SIMILARITY.
FT	DISULFID	2070	2097	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	333	333	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	387	387	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	611	611	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	667	667	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1842	1842	N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE	2124 AA;	221117 MW;	E30BBE61593A34B1 CRC64;

Query Match 4.6%; Score 235.5; DB 1; Length 2124;  
 Best Local Similarity 21.3%; Pred. No. 0.0049; Mismatches 359; Indels 353; Gaps 48;  
 Matches 224; Conservative 115;

Qy	112	NGVWSIAADPINTNKVAAVGMVNTNSWDPNDGAILRSSDQAGATWQITPLPFLKGMNPGR	171
Db	979	SGIEDISVLPTGESPETASGVGLSLPGSGESLETASGVY-DVTQLPTERGGLTSA	1037
Qy	172	GMGERLAVDPNNNDNL-----YFGAPSGKGLWRSTDGATWSQMTNFPD-----	215
Db	1038	SGIEDITVLPTGRENLTSASGVEDVSLPGSGEGLETASGI--EDISVFTEAEGLT	1095
Qy	216	--VGTIAN-PTDTTGYQSDIQGVVWVAFDRSSSISGQASKTI-FGVADPNPNPVFWSRDG	272
Db	1096	SASGYVSGIPSGEDGTETSTSGVEGVS---GLPSGEGLETSGSGVEDLGLPTRDSLET	1152
Qy	273	GATWQAVPGAPTG-----FIPHKGVDPVNVHLY-----IATNTG-----GPLYD-	312
Db	1153	SASGVDTVTGYPGREDTETSP--GVGDDLSLPGSQEGLETASGAEDLGLPSCKEDL	1210
Qy	313	-GSSGDVWKF-----SVTSGTWTRISVPSPDTANDYFG-----YSLG----	349
Db	1211	VGSAGALDFGKLPSTGLSGQTPEASGLPS-GFSGEYSGVDIGSGPSPGLPDSGLPSG	1269
Qy	350	--TDRQHPNTIMVAT-----QISWVPTIIFRSTD--GGATWTRI	386
Db	1270	FFTVSLVDSTLVEVITATTASELEGRGTVSVSGSESGEPPLSELDSSADISGLPSGTEL	1329
Qy	387	WDWTSYPNRSLRYVLDISAEPWLTGQVQNP-----PVPSKLGWM	427
Db	1330	SGQTSGS-----LDVSGETSGFFDVSGQPPGSGTGEGTSGIPEVSQAVRSPPTTEI	1382
Qy	428	DEAMADIPFNSD-----RMLYGTGATLYATNDLTKWDSGGQIHIAPMVKGLEETAVNDL	481
Db	1383	SELSCLSGQPDVSGEGSGLFGSQSGSITS-----VSG-ETSGISDL	1425
Qy	482	ISPSGAPLISAL-----GDLOGFTHADVAVPSTIFTSFVFTTGTSDY	526
Db	1426	SCQPSGFPVLSGTTPTGTPDLASGAMSGSGDSGITFVD-----TSLIEVTPPTFRE	1476

```
QY 527 ABLNPSLIVRAGSFDPSQNDHRHAFSTGGKWFQSGSRPGVTTGTTVAASADGRFV 586
Db 1477 EE-----GLGSVELSGLPSGETLSTGTMVD-VSGSSGAIDSSGLISPTPEFS--- 1525
QY 587 WAPCDPQPVVYAVFGNSWAAAGVAFAN-AQIRSDRVNPKTFYALNGNGFYRSTDG--- 642
Db 1526 -----GLPSGVAEVSGEVSGVETGSSLSGAP-----DGSGL 1557
QY 643 -----GVTFQVPA--AG-----LPSSGAVGVMFHVPKRGDL----- 673
Db 1558 VSGPPTVSLVDRITLVESITLAPTAQEAAGEGSSILEFSGA-----HSGTPDISGLSGSLD 1613
QY 674 -----WLAAS-----SGLYHSTN--GGSSWGAITGVSAVNVFGKSAAGSSYP 715
Db 1614 QSTWQPGWTEASTEPSSPFSGDFSTTDASGESITAPTG-----SGETSGLP 1662
QY 716 AVFVGT--IGVGTGAYRSDCGTTWVLINDQHOYGNWGAITGDHANLRVVICNGR 773
Db 1663 EVTLITSELVEGVTEPTVSOELG-----HGFSMT-----YTRLPF-EASGE 1702
QY 774 GIVYDGG-----APSGSPSPSVSPASPSLSPSPSSSPSPSPSSSP 820
Db 1703 ASAGDLGGPVTIPPGSGVSAPEGSDFSAYPEAGVGVSAAPSAEQSLSEFPDLHGIT 1762
QY 821 SSS-----PSPSPSPSPSP-----SRSPSPSPSPSPSPSPSPSPSPSPSPSP 869
Db 1763 SASRETLEMTTPTGTEVSSNFWTFQETREGSAAPEVSGSSSTSDIDAGTSGVPPATPM 1822
QY 870 SGGVKVQYKNDAPGDNQIKPLQVNTGSSSVDLSTVT--VRYWTRDGGSSSTLYNC 927
Db 1823 TSGDRTEISGEWD-----HTSEVNVTVSTVPESRWAQSTQHPETL----- 1865
QY 928 DWAAIGCGNIRASFGSNPA-----TPTADT 953
Db 1866 -----QEIGSPNPSYSGEETQTAET 1885
```

## RESULT 13

```
VTX TPX TTV1 STANDARD; PRT; 360 AA.
ID _VTX TPX TTV1 STANDARD; PRT; 360 AA.
AC P19274;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Viral protein TPX.
OS Thermoproteus tenax virus 1 (strain KRAL) (TTV1).
OC Viruses; dsDNA viruses, no RNA stage; Lipothrixviridae;
OC Lipothrixvirus.
OX NCBI_TaxID=10480;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90174928; PubMed=2308830;
RA Neumann H., Zillig W.;
RA "The TTV1-encoded viral protein TPX: primary structure of the gene
RT and the protein.";
RL Nucleic Acids Res. 18:195-195(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X14855; CAA33002.1; -.
CC PIR; S12850; S12850.
KW Repeat.
FT DOMAIN 278 353 3 THR-PRO REPEATS REGIONS AND TWO NEAR
FT REPEAT 270 291 IDENTICAL REPEATS.
FT REPEAT 292 301 THR-PRO (N).
FT REPEAT 302 322 THR-PRO (N).
```

```
FT REPEAT 323 332 THR-PRO (N).
FT REPEAT 333 353
SQ SEQUENCE 360 AA; 38214 MW; EF54PBCB753EDICA CRC64;
Query Match 4.3%; Score 222.5; DB 1; Length 360;
Best Local Similarity 22.5%; Pred. No. 0.0027;
Matches 83; Conservative 55; Mismatches 94; Indels 137; Gaps 17;
QY 581 DGSRFVW-----APGDGPQPVVYAVFGNSWAAAGVAFAN-AQIRSDRVNPKTFYALSNGT 635
Db 48 DG---VWVVKLRSLSPGQSITI-----TASSGTP-----NIDPTIALYNNGS 88
QY 636 FYRSTD-----GGVTFQVPAAG-----LPSSGAVGVMFHVPKRG 671
Db 89 SYSNLTIVGSPVTIVQDFGGVYSAISAYASGDFLVAASPTGFTPSSRLLV----- 139
QY 672 DLWL-----AASGLYHSTNGSSWSAITG--VSSAVNVFGKSAAGSSYPVAVFVGT 722
Db 140 DRWATPTSLDAVGLRLYADTN--DWFGVVRKYINGAQNVSLEQKISG----- 185
QY 723 IGVGTGAYRSDCGTTWVLIND-QHOYGNWGAITGDHANLRVVICNGRGI 775
Db 186 -----TYSVNEIDISQFAAFTDPLVWYLSINGSTANV-KVYKQGNIGT 229
QY 776 VYGDIGGAPSGSPS-----PSVSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS 815
Db 230 VSGNVSTTPYGNPSPMAGYGTVDKHYANFIVLPYEPDQVTPVPISSPSPSTPTPTPTPT 289
QY 816 PSSSPS---SPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 861
Db 290 PTYDITYVVFVOTSPSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 349
QY 862 PTPSSSPVS 870
Db 350 PTPSTTSS 358
RESULT 14
MUSB HUMAN
ID _MUSB HUMAN STANDARD; PRT; 5703 AA.
AC Q9HC84; Q00447; Q00573; Q14985; Q15494; Q95291; Q95451; Q14881;
AC Q99552; Q9UE28;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mucin 5B precursor (Mucin 5 subtype B, tracheobronchial) (High
DE molecular weight salivary mucin MGI) (Sublingual gland mucin).
GN MUC5B OR MUC5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-1594 FROM N.A.
RA Chen Y., Di Y.P., Wu R.;
RT "Molecular cloning of the amino-terminal and 5'-flanking region of the
RT human MUC5B mucin gene.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-1325 FROM N.A.
RX MEDLINE=99009274; PubMed=9790959;
RA Offner G.D., Nunes D.P., Keates A.C., Afdhal N.H., Troxler R.F.;
RT "The amino-terminal sequence of MUC5B contains conserved
RT multifunctional D domains: implications for tissue-specific mucin
RT functions.";
RL Biochem. Biophys. Res. Commun. 251:350-355(1998).
RN [3]
RP SEQUENCE OF 40-1324 FROM N.A.
RX MEDLINE=99023932; PubMed=9804771;
RA Desseyn J.-L., Buisson M.P., Porchet N., Aubert J.-P., Laine A.;
RT "Genomic organization of the human mucin gene MUC5B: cDNA and genomic
RT sequences upstream of the large central exon.";
RL J. Biol. Chem. 273:30157-30164(1998).
```

[4] SEQUENCE OF 1326-4895 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=97166151; PubMed=9013550;  
RA Desseyn J.-L., Guyonnet-Duperat V., Porchet N., Aubert J.-P.,  
Laine A.;  
RT "Human mucin gene MUC5B, the 10.7 kb large central exon encodes  
various alternate subdomains resulting in a super-repeat. Structural  
evidence for a lip15.5 gene family.";  
RL J. Biol. Chem. 272:3168-3178 (1997).  
[5] SEQUENCE OF 4057-4480 FROM N.A.  
RC TISSUE=Salivary gland;  
RX MEDLINE=97292540; PubMed=9147051;  
RA Nielsen P.A., Bennett E.P., Wandall H.H., Therkildsen M.H.,  
Hannibal J., Clausen H.;  
RT "Identification of a major human high molecular weight salivary mucin  
(MG1) as tracheobronchial mucin MUC5B.";  
RL Glycobiology 7:413-419 (1997).  
[6] SEQUENCE OF 4721-5703 FROM N.A.  
RC TISSUE=Gall bladder;  
RX MEDLINE=97293229; PubMed=9164870;  
RA Keates A.C., Nunes D.P., Afzal N.H., Troxler R.F., Offner G.D.;  
RT "Molecular cloning of a major human gall bladder mucin: complete C-  
terminal sequence and genomic organization of MUC5B.";  
RL Biochem. J. 324:295-303 (1997).  
[7] SEQUENCE OF 4809-5687 FROM N.A.  
RC TISSUE=Sublingual gland;  
RX MEDLINE=96125355; PubMed=8554565;  
RA Troxler R.F., Offner G.D., Zhang F., Iontcheva I., Oppenheim F.G.;  
RT "Molecular cloning of a novel high molecular weight mucin (MG1)  
from human sublingual gland.";  
RL Biochem. Biophys. Res. Commun. 217:1112-1119 (1995).  
[8] SEQUENCE OF 4859-5703 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=97347489; PubMed=9201995;  
RA Desseyn J.-L., Aubert J.-P., Porchet N., Laine A.;  
RT "Genomic organization of the 3 region of the human MUC5B mucin.";  
RL J. Biol. Chem. 272:16873-16883 (1997).  
CC -1- FUNCTION: Salivary mucin that is thought to contribute to the  
lubricating and viscoelastic properties of whole saliva.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- TISSUE SPECIFICITY: Expressed mainly in bronchus glands and also  
in submaxillary glands, endocervix, gall bladder, and pancreas.  
CC -1- PTM: Highly glycosylated.  
CC -1- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.  
CC -1- SIMILARITY: Contains 3 VWFC domains.  
CC -1- SIMILARITY: Contains 4 VWFC domains.  
CC -1- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF107890; AAC33673.1; -;  
EMBL; AF066604; AAC67545.1; -;  
EMBL; AJ004862; CAA06167.1; -;  
EMBL; Z72496; CAA96577.1; -;  
EMBL; X74955; CAA52910.1; -;  
EMBL; U63836; AAB61398.1; -;  
EMBL; U78554; AAC51344.1; -;  
EMBL; U78552; AAC51344.1; JOINED.  
EMBL; U78553; AAC51344.1; JOINED.  
EMBL; U78551; AAC51343.1; -;  
EMBL; U95031; AAB65151.1; -;  
EMBL; Y09788; CAA70926.1; -;

DR Genew; HGNC:7516; MUC5B.  
DR MIM; 600770; -;  
DR GO; GO:0005515; F:protein binding; IPI.  
DR InterPro; IPR006208; Cys\_knot.  
DR InterPro; IPR006207; Cys\_knot\_C.  
DR InterPro; IPR009041; PMP\_inhibitor.  
DR InterPro; IPR002919; TIL\_Cysrich.  
DR InterPro; IPR006552; VC\_out.  
DR InterPro; IPR001007; VWFC\_C.  
DR InterPro; IPR001846; VWFC\_D.  
DR Pfam; PF00007; Cys\_knot; 1.  
DR Pfam; PF01826; TIL; 1.  
DR Pfam; PF00093; vwc; 1.  
DR Pfam; PF00094; vwd; 4.  
DR SMART; SM00214; VWC; 6.  
DR SMART; SM00215; VWC\_out; 4.  
DR SMART; SM00216; VWD; 4.  
DR PROSITE; PS01185; CTCK\_1; 1.  
DR PROSITE; PS01225; CTCK\_2; 1.  
DR PROSITE; PS01208; VWFC\_1; 2.  
DR PROSITE; PS01184; VWFC\_2; 2.  
DR Glycoprotein; Repeat; Signal; Polymorphism.  
FT SIGNAL 1 25  
FT CHAIN 26 5703  
FT DOMAIN 77 225  
FT DOMAIN 329 386  
FT DOMAIN 426 580  
FT DOMAIN 858 918  
FT DOMAIN 896 1044  
FT DOMAIN 1457 1603  
FT DOMAIN 1609 4873  
FT DOMAIN 5005 5178  
FT DOMAIN 5353 5425  
FT DOMAIN 5462 5528  
FT DOMAIN 5594 5683  
FT DISULFID 5594 5646  
FT DISULFID 5622 5660  
FT DISULFID 5626 5676  
FT DISULFID 5643 5678  
FT DISULFID 5645 5682  
FT CARBOHYD 145 145  
FT CARBOHYD 201 201  
FT CARBOHYD 254 254  
FT CARBOHYD 402 402  
FT CARBOHYD 516 516  
FT CARBOHYD 806 806  
FT CARBOHYD 930 930  
FT CARBOHYD 1277 1277  
FT CARBOHYD 1293 1293  
FT CARBOHYD 1557 1557  
FT CARBOHYD 1775 1775  
FT CARBOHYD 2192 2192  
FT CARBOHYD 2721 2721  
FT CARBOHYD 3419 3419  
FT CARBOHYD 3948 3948  
FT CARBOHYD 4745 4745  
FT CARBOHYD 4901 4901  
FT CARBOHYD 4958 4958  
FT CARBOHYD 4965 4965  
FT CARBOHYD 4987 4987  
FT CARBOHYD 5037 5037  
FT CARBOHYD 5052 5052  
FT CARBOHYD 5156 5156  
FT CARBOHYD 5427 5427  
FT CARBOHYD 5467 5467  
FT CARBOHYD 5506 5506  
FT CARBOHYD 5507 5507  
FT CARBOHYD 5543 5543  
FT CARBOHYD 5553 5553  
FT CARBOHYD 5604 5604  
FT CARBOHYD 5618 5618  
FT CARBOHYD 5662 5662  
FT CARBOHYD 5662 5662  
FT VARIANT 5137 5137  
T -> S (in dbSNP:2672788).





the protein (mutation in the G1 domain).

-1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.

-1- SIMILARITY: Contains 4 link domains.

-1- SIMILARITY: Contains 1 EGF-like domain.

-1- SIMILARITY: Contains 1 C-type lectin family domain.

-1- SIMILARITY: Contains 1 Sushi (SCR) domain.

-1- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

EMBL; LO7049; AAC37670.1; --  
EMBL; S73722; AAB32160.1; --  
DR EMBL; S73721; AAB32160.1; JOINED.  
DR PIR; A55182; A55182.  
DR HSSP; P98066; 1TSG.  
DR MGD; MGI:99602; Agcl.  
DR InterPro; IPR002353; Antifreeze2.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003006; Ig-MHC.  
DR InterPro; IPR003596; Ig V.  
DR InterPro; IPR001304; Lectin\_C.  
DR InterPro; IPR000538; Link.  
DR InterPro; IPR003324; SGXXSG.  
DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
DR Pfam; PF00047; Ig; 1.  
DR Pfam; PF00059; lectin\_C; 1.  
DR Pfam; PF02339; SGXXSG; 60.  
DR Pfam; PF00084; sushi; 1.  
DR Pfam; PF00193; Xlink; 4.  
DR PRINTS; PR00356; ANTIFREEZEII.  
DR PRINTS; PR01265; LINKMODULE.  
DR ProDom; PD000918; Link; 4.  
DR SMART; SM00032; CCP; 1.  
DR SMART; SM00034; CLECT; 1.  
DR SMART; SM00406; IG; 1.  
DR SMART; SM00445; LINK; 4.  
DR PROSITE; PS00615; C TYPE LECTIN\_1; 1.  
DR PROSITE; PS50041; C TYPE LECTIN\_2; 1.  
DR PROSITE; PS50835; IG LIKE; 1.  
DR PROSITE; PS00290; IG\_MHC; 1.  
DR PROSITE; PS01241; LINK; 4.  
KW Glycoprotein; Proteoglycan; Lectin; Signal; Sushi; Repeat;  
KW Immunoglobulin domain.

SIGNAL 1 19  
FT CHAIN 20 2132  
FT DOMAIN 34 147  
FT DOMAIN 170 247  
FT DOMAIN 268 349  
FT DOMAIN 504 581  
FT DOMAIN 602 683  
FT DOMAIN 1918 2044  
FT DOMAIN 2048 2106  
FT DOMAIN 48 140  
FT DOMAIN 152 247  
FT DOMAIN 253 349  
FT DOMAIN 486 580  
FT DOMAIN 587 682  
FT DOMAIN 685 803  
FT DOMAIN 805 1231  
FT DOMAIN 1232 1917  
FT DOMAIN 1917 2132  
FT DOMAIN 51 133  
FT DISULFID 175 246  
FT DISULFID 199 220  
FT DISULFID 273 348  
FT DISULFID 297 318  
FT DISULFID 509 580

POTENTIAL.  
AGGRECAN CORE PROTEIN.  
IG-LIKE V-TYPE.  
LINK 1.  
LINK 2.  
LINK 3.  
LINK 4.  
C-TYPE LECTIN.  
SUSHI.  
G1-A.  
G1-B.  
G1-B'.  
G2-B.  
G2-B'.  
KS.  
CS-1.  
CS-2.  
G3.  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.

FT	DISULFID	533	554	BY SIMILARITY.
FT	DISULFID	607	682	BY SIMILARITY.
FT	DISULFID	631	652	BY SIMILARITY.
FT	DISULFID	1922	1933	BY SIMILARITY.
FT	DISULFID	1950	2042	BY SIMILARITY.
FT	DISULFID	2018	2034	BY SIMILARITY.
FT	DISULFID	2049	2092	BY SIMILARITY.
FT	DISULFID	2078	2105	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	333	333	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	387	387	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	611	611	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	667	667	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1675	1675	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	SITE	1171	1173	CELL ATTACHMENT SITE (POTENTIAL).
SQ	SEQUENCE	2132 AA;	222008 MW;	OB2BCDFC6CDBA163 CRC64;

Query Match 4.3%; Score 221; DB 1; Length 2132;  
Best Local Similarity 20.9%; Pred. No. 0.021;  
Matches 199; Conservative 102; Mismatches 311; Indels 338; Gaps 46;

Qy	110	GYNGVWSIAADPINTNKVAAVGMVYNSWDNDGAILRSSDOGA-----TWQITPL	160
Db	1105	GYVSGIPSGDGTET-----SASGVEDVSGLPGGEGLETSASGVDELGPSTRDSLETSAS	1160
Qy	161	PFKLGNNMGRGMRERLAVDPNNDNILYFGAPSGKGLWRSTDGATWSQMTNPPD-----	215
Db	1161	GVDVTGPGSGRDPETSVSGVGD-----FSGLPKGEGLETSASGA--EDLSGLPSGKEDL	1215
Qy	216	-----VGTIANPT-DTTQYQSDIQGVVWVAFPKSSSLGQAKTIFFGVA	260
Db	1216	VGSASGALDFGKLPGLTGLSGQTPEVNGFPSPGSG-----EYSGADISGSPSS---GLP	1266
Qy	261	DPNNPVFWSRDGGATQAVPGAGTGFPHKGVDFPNHVLXIATSNVTGPGYDG-----	313
Db	1267	D-----FSGLPSPGF-PTVSLVD--STLVEVITATTSLEGRGTIGIS	1306
Qy	314	SSGDVWKFVTSQWTRISPVSTANDYFGYSGLTIDRHPNTIMVATQISWPDITII	373
Db	1307	CSGEV-----SGL-----PLGELDSSAD--ISGL-----PSGTLSQASGSPDS--	1344
Qy	374	FRSTDGGATWTRIMDWTSPNRSRLYVLDISAPMLTTFGVQNPVPSPKLGMDAMAI	433
Db	1345	SGETSG-----PFDVSGQPFSSGVS-----EETSGI	1371
Qy	434	DFNSDRMLYGTGATLYATNDLTKWD--SGGQIHIAPMVKGL-----EET	476
Db	1372	PEISGQPS--GTPDPTT-ATSGVTELNELSSGQPDVSGDGSGLFGSGQSGITSVSGETS	1428
Qy	477	AVNDLISPSPGAPL-----ISALDGLGGFTHADVTAVPSTIFT-----	514
Db	1429	GISLDSGQSPGFPVPSGTATRTPDLASGTISGSGSSGITFVDTSFVEVPTTFREEGEGL	1488
Qy	515	-----SPVFTTCTSDVYAEALNPSTIIRAGSFDPPSQPNDRHVAFTDGGKNWF	562
Db	1489	GSVELSGFPFGSETLSGTSGTVDVSEQSSGAISSGLTSPTE-----F	1532
Qy	563	QGSBPG-----GVTGTGTVAAASA-DGSRFVWAPGDP-----GQPVVY	598
Db	1533	SGLPSGVAEVSQGEFSGVETGSSLPSCAFDGSGLV--SGFPTVSLVDRTLVESTQAPTAQ	1590
Qy	599	AVRGNSW-----AASQGVPAQAQRSDRVNPKTFYALNSGTFYRSTDGGVTQPVAAGL	653
Db	1591	EAGGPGSGLIFSGAHSGTP---DISGELSGSLDLSLTSQGMETSTE-----T	1636
Qy	654	PSSGAVGVWFHAPVCKEGDMLAASSGLYHSTNGSSWSAITGVSSAVNVFGKSPGSS	713
Db	1637	PSS-----PYFSGDF-----SSTTDVSGESIAATTGSGES---SG	1668
Qy	714	YPAVFVWGT--IGGVGTGAYRSDCCGTTWVLIINDQHQYGNWGAITGDHANLARVIG--	769
Db	1669	LPEVTINTSELVEGVTETVTSQELG-----HGFSMT-----YISRL	1704



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2004, 09:13:38 ; Search time 49 Seconds  
(without alignments)  
6162.264 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MDRSENIRLTMRSRLVSL.....RASFGSVNPATPTADTVLQX 957

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2479.5	48.3	882	16 Q82M04	Q82m04 streptomyc
2	2436	47.4	890	16 Q86727	Q86727 streptomyc
3	2420.5	47.1	996	2 Q9A0H0	Q9a0h0 caldicellul
4	2037	39.7	839	16 Q97KK0	Q97kk0 clostridium
5	1831.5	35.7	857	3 Q8TFP1	Q8tfp1 aspergillus
6	1784	34.7	806	3 Q9P4T8	Q9p4t8 agaricus bi
7	1740	33.9	856	3 Q74170	Q74170 aspergillus
8	1625.5	31.7	838	3 Q7Z9M8	Q7z9m8 trichoderma
9	1442	28.1	739	16 Q82K30	Q82k30 streptomyc
10	1173.5	22.9	751	16 Q8P1M5	Q8p1m5 xanthomonas
11	1161	22.6	707	16 Q9WTE1	Q9wte1 thermocoga
12	1151.5	22.4	751	16 Q8P9U5	Q8p9u5 xanthomonas
13	1125	21.9	812	3 Q8J0D2	Q8j0d2 geotrichum
14	431.5	8.4	1779	2 Q52374	Q52374 caldicellul
15	409	8.0	196	3 Q8J1H7	Q8j1h7 agaricus bi
16	409	8.0	1770	2 Q9X3P5	Q9x3p5 caldicellul

17	402.5	7.8	921	2 Q9L8L8	Q9l8l8 caldibacill
18	398.5	7.8	1711	2 P96311	P96311 anaerocellu
19	387	7.5	2014	16 Q7U3X4	Q7u3x4 synecococc
20	385	7.5	1064	2 Q7X2U2	Q7x2u2 uncultured
21	372.5	7.3	473	10 Q39620	Q39620 chlamydomon
22	350.5	6.8	997	2 Q9Z411	Q9z4i1 bacillus sp
23	346.5	6.7	901	2 Q44562	Q44562 actinomyc
24	336.5	6.6	741	16 Q82QP2	Q82qf2 streptomyc
25	334	6.5	1751	2 Q9A0G4	Q9a0g4 caldicellul
26	325.5	6.3	930	2 Q9RFX5	Q9rfx5 caldibacill
27	325	6.3	1000	2 Q24820	Q24820 thermophili
28	320.5	6.2	991	10 Q94C44	Q94c44 chlamydomon
29	318.5	6.2	875	2 Q9F2B0	Q9f2b0 thioobacillu
30	316	6.2	1091	2 Q8KKF7	Q8kkf7 paenibacill
31	308.5	6.0	611	16 P74375	P74375 synecocyst
32	301.5	5.9	1749	16 P73032	P73032 synecocyst
33	298.5	5.8	1154	16 Q7U5X7	Q7u5x7 synecococc
34	297	5.8	714	16 Q7U5X6	Q7u5x6 synecococc
35	297	5.8	1915	2 Q9RPL0	Q9rpl0 acetivibrio
36	294	5.7	474	5 Q86I54	Q86i54 dictyosteli
37	292	5.7	616	2 Q7X2N2	Q7x2n2 thermomonos
38	291	5.7	351	10 Q39492	Q39492 chlamydomon
39	290.5	5.7	2232	5 Q8IFX6	Q8ifx6 caenorhabdi
40	288.5	5.6	1032	5 P91365	P91365 caenorhabdi
41	286	5.6	261	2 Q9AQG7	Q9a0g7 caldicellul
42	279.5	5.4	825	5 Q7YZW4	Q7yzw4 caenorhabdi
43	279	5.4	170	2 Q9RFX6	Q9rfx6 caldibacill
44	277	5.4	234	12 Q91GH4	Q91gh4 epiphyas po
45	276	5.4	1426	2 Q9X3P6	Q9x3p6 caldicellul

## ALIGNMENTS

### RESULT 1

Q82M04	ID	Q82M04	PRELIMINARY;	PRT;	882 AA.
AC	Q82M04;				
DT	01-JUN-2003	(TREMBLrel. 24, Created)			
DT	01-JUN-2003	(TREMBLrel. 24, Last sequence update)			
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)			
DE	Putative endo-1,4-beta-glucanase.				
GN	CELA3 OR SAV1856.				
OS	Streptomyces avermitilis.				
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
OC	Streptomycineae; Streptomycetaceae; Streptomycetes.				
OX	NCBI_TaxID=33903;				
RP	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;				
RX	MEDLINE=21477403; PubMed=11572948;				
RA	Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,				
RA	Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,				
RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.;				
RT	"Genome sequence of an industrial microorganism Streptomyces				
RT	avermitilis: deducing the ability of producing secondary				
RT	metabolites."				
RL	Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).				
RL	[2]				
RN	SEQUENCE FROM N.A.				
RP	STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;				
RC	MEDLINE=22608306; PubMed=12692562;				
RX	Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,				
RA	Sakaki Y., Hattori M., Omura S.;				
RT	"Complete genome sequence and comparative analysis of the industrial				
RT	microorganism Streptomyces avermitilis."				
RL	Nat. Biotechnol. 21:526-531(2003).				
DR	EMBL; AP005028; BAC69567.1; --				
DR	GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.				
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.				
DR	InterPro; IPR001919; Bac cellose-bind.				
DR	InterPro; IPR008965; Cellul bind.				
DR	InterPro; IPR002860; GH_BNR_				

DR Pfam; PF02012; BNR; 10.  
DR Pfam; PF00553; CBM\_2; 1.  
DR SMART; SM00637; CBD\_II; 1.  
KW Complete proteome.  
SQ SEQUENCE 882 AA; 92152 MW; A3C8E753CB69F13D CRC64;

Query Match 48.3%; Score 2479.5; DB 16; Length 882;  
Best Local Similarity 50.1%; Pred. No. 1.7e-109;  
Matches 476; Conservative 129; Mismatches 230; Indels 115; Gaps 15;

QY 11 MRSRLVSLAATASFAVAAL--GVLPITAITASPAHAATTPQVTSNVAIGGGFVDGI 68  
DB 1 MRRTRIF-----TAVLALAGLPAGTPPALAAPTATIAADTYWKNARVDGGFVPGI 55  
QY 69 VFNEGAPGILYVRTDIGGMYRDAANGRWIPLLDWGNWNNWYGVVSYAADPINTKMW 128  
DB 56 VFNSEKNLAYRTDIGGARMAESSKTWPLDLSVGSMDGHTGVVSLASDSVDPNKVY 115  
QY 129 AAVGMTNSWDNDGAILRSDGATWQITPLPFKLGNNPGRGMGERLAVDNNDNLY 188  
DB 116 AAVGTYTNSWDNPGAVLRSGDRGASQKTDLPFKLGNPGRGMGERLAVDNRSVLY 175  
QY 189 FGAPSGKGLWRSDSGATWQMTNFPDVGTYIANPTDTTCYQSDIQGVVWVAFDKSSSL 248  
DB 176 LGAPSGKGLWRSDSGASWSQVTDFFNVGYVQDATDTSGYASDNQGVVWVTFDESGSP 235  
QY 249 GQASKTIFGVADPNPNPFFVWSRDGGATWQAVGAPTGFIPHKGVDFPNVHVLVIATNTG 308  
DB 236 GSSTRIVYGVADKNSVYSTDAGATWSLGAQPTGHLAKGVLDAAAGCLVLYASDKG 295  
QY 309 GPVDSGSDVWKFSTVGTWTRISPVSTDTANDYFGYGLTIDRQHPNTIMVATQISWW 368  
DB 296 GPYDGGKGLWRVTTKTGTNTISPAVAEDT---YGFSGLTVDHQHPGTWATAYSSWW 352  
QY 369 PDTIIFRSTGGATWTRIMWTSPYRSLRYVLDDISAEPLWTFGVQPNPVPSPKLGMD 428  
DB 353 PDTQLFRSTDSGGTWTWKAMDYTSYPKSNRFTMDVSSPWLWTGANPAPPEQTPKLGWMT 412  
QY 429 EAMAIDPFNSDRMLYGTATLYATNDLTWKDGGQHIAPMWGLBETA VNDLISPPS-G 487  
DB 413 ESLIEDPFDARMYGTATVYGTDLNTWDSQSQFTIKPMARGLEBETA VNDLASPPSGG 472  
QY 488 APLISALDGLGFTHADVTAVPSTIFTSVFTTCTSDYVLAELNPSIIVRAGSFDPPSQPN 547  
DB 473 AQLFSAIDGLGFRHTDLTTPSLMITSNPTFTSTSLDYAETDPGTVVRYGNLD--SGP- 529  
QY 548 DRVIAFSTGGKWFQSGEGVTGTGTAASADGSRFVWAPDGPQGVVYVGFPGNSWA 607  
DB 530 --HVAFTDNGANWFAGADPSVSGGTVAASDGRFVWSPAGTG--VQYTTGFGTWS 585  
QY 608 ASOGVPANAQIRSDRVNPKTFYALSNGTFYRSTGDTGTFQPVAA--GLPSSGAVGVMFHAV 666  
DB 586 ASAGLPAGALVESDRVDPKTFYFKSGRFVYSSDGGATFTASATGLPSSGDS--VRFKAL 643  
QY 667 PGKEGLWLAASS-----GLYHSTNGSSWSAITGVSSAVNVFGFKSAPGSSYPAVFVVG 721  
DB 644 PGTKGIWLAGASDAGYGLWHSTDGGAAFTKLATVDQATIGFGKAATGASQYTLTSA 703  
QY 722 TIGGVTAGRSDDCGTTWVLINDHQYGNWQAITGDHANLARVYIGTNGRGIVYGDIG 781  
DB 704 KIGGVRGIFRSTDKGASWTRVNDHAWGWTGAITGDPRVYGVVYSTNGRGIVYGDTA 763  
QY 782 GAPSGSPSPVSPASPSLSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS 841  
DB 764 GSSDGG----- 769  
QY 842 ASP 901  
DB 770 -----GTEAPT-----GACTVYTRITQWGGFQ--ADVQLANTGST 805  
QY 902 SVDLSTVTVRYWTRDGGSSSTLVNCDWAAIGCG-----NIRASFGS 943  
DB 806 AWDGWSLIG---WSFGDQGEVTLQMNASYAQAGSGVTAANLAWNGRVAAGS 852

RESULT 2  
O86727 PRELIMINARY; PRT; 890 AA.  
AC O86727;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Putative secreted cellulase.  
GN SCO6545 OR SC5C7.30C.  
OS Streptomyces coelicolor.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Seeger K.J., Harris D.;  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Parkhill J., Barrell B.G., Rastam M.A.;  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
MEDLINE=21996410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
Harper D., Batenan A., Brown S., Chandra G., Chen C.W., Collins M.,  
Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
Huang C.-H., Kieser T., Lark L., Murphy L., Oliver K., O'Neill S.,  
Rabinowitz K.E., Rastam M.A., Rutherford K., Rutter S.,  
Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
DR EMBL; AL939128; CAA20642.1; --  
DR PIR; T35237; T35237.  
DR HSP; P07986; IEXG.  
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR001919; Bac celose-bind.  
DR InterPro; IPR008965; Cellul\_bind.  
DR InterPro; IPR002860; GH\_BNR.  
DR Pfam; PF02012; BNR; 10.  
DR Pfam; PF00553; CBM\_2; 1.  
DR SMART; SM00637; CBD\_II; 1.  
KW Complete proteome.  
SQ SEQUENCE 890 AA; 93252 MW; D2C58695A4B56E84 CRC64;

Query Match 47.4%; Score 2436; DB 16; Length 890;  
Best Local Similarity 48.6%; Pred. No. 1.9e-107;  
Matches 475; Conservative 129; Mismatches 246; Indels 126; Gaps 19;

QY 11 MRSRLVSLAATASFAVAALGVLPITAITASPA--HAATTPQVTSNVAIGGGFVDGI 68  
DB 1 MRRTRITVLLALAAGLIA---GSPFAASAAEPAPRAVAADSYTKNARIDGGGFVPGI 57  
QY 69 VFNEGAPGILYVRTDIGGMYRDAANGRWIPLLDWGNWNNWYGVVSYAADPINTKMW 128



```

QY 666 VPKEGDLMLAASSGLYHSTNGSSWSAITGVSSAVNVGFGKAPGSSYPVAFVVGTTGG 725
DB 646 VPIEGEDIMLVGNNGMWSHSDGGSYFVKISGVEDAASIGFKPAEGETYPALITYAKING 705
QY 726 VTGAYSDDCGTTWVLINDDOHQYGNWQAITGDHANLRVYIGTNGRGIVYDGGAPS 785
DB 706 VRGIFRSDCDKTIWIRINDDKHOFGCANADITGDPVYGVFVATNGLGKWEIA--Y 762
QY 786 GSPSPVSPASPSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 845
DB 763 SNISPSATPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPSTPST 822
QY 846 PSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 904
DB 823 MTPTPTP-----TPTPTPTGPTGGLKVLKNNETSASAGSIRWFVKVNGSSSVD 876
QY 905 LSVTVRYWFTRGGSSTLYNCWAAAGCNRASFGSNVATPTADTYLO 956
DB 877 LSRVKIRYWTVDGKPSQSAV-CDWAQIGASNTVFNFVGLSSGVSGADYYLE 927

RESULT 4
Q97KK0 PRELIMINARY; PRT; 839 AA.
AC Q97KK0;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probably secreted sialidase, several ASP-boxes and dockerin
DE domain.
GN CAC0919.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Berton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hatti J., Wolf Y.I.,
RA Tatusov R.N., Sabathe F., Doucet-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007608; AAK78895.1; -.
DR PIR; D97013; D97013.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR02105; Dockerin_1.
DR InterPro; IPR02048; EF-hand.
DR InterPro; IPR02860; GH_BNR.
DR Pfam; PF02012; BNR; 9.
DR Pfam; PF00404; Dockerin_1; 2.
DR PROSITE; PS00018; EF_HAND; 2.
KW Complete proteome.
SQ SEQUENCE 839 AA; 90824 MW; 6C99A041CA9CF984 CRC64;

Query Match 39.7%; Score 2037; DB 16; Length 839;
Best Local Similarity 49.5%; Pred. No. 1.3e-88;
Matches 392; Conservative 118; Mismatches 252; Indels 30; Gaps 16;

QY 13 SRLVSLAA-----TASFVAALGVLPIAITASPAHAAT-TQPYTWSNVVIGGGGFVD 66
DB 2 NKRIVSMVAGLSIIFTGF-----VTHISAANKAAASVQSQYKWDNAKIGAGGYVP 54
QY 67 GIYVNEGAPILYVRTDIGMYRWDAANGRWIPLLDWVGNWNGYGVVIAADPIYTNK 126
DB 55 AVIFNTEFODLIYARTDMGAYRWKANKWIPITD--GFSWTMLGCEIATDPIDTNR 112
QY 127 VWAAGVMTYNSWPNDCGAILRSSDQAGATQITPLPFKLGGMPCRGWGERLAVDPNDNI 186

```

```

DB 113 VYIAGYITNDQDENAVILSSQKGNWTKRYQLPFFKVGNNPGRNMGRQLQIDPNDDKI 172
QY 187 LYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDTTGYQSDIQGVVWVAFDKSS 246
DB 173 LYLARSNGLMKSEDIYQOTWSKVDNFPDGTGYVDQPQNE--YTADKVGVMWETFDPSIG 230
QY 247 SLGQASKTIYFVGAD-PNNVFWMSRDGGATWQAVPGAPGTGTFPHKGVFDFVNVHLVIAT 305
DB 231 TKGSPTQTMVYGAADKTGNNIYVTNDGKTKWSAVKQPKGYLPHHGIL-ASDGMLYISYS 289
QY 306 NTGSPYDSSGDVWKFSTGWTIRISVPSTDDTANDYFGYSLGITDRDHPNTIMATQI 365
DB 290 NTCFYPDGSQGVWKNYTKTGEWNTITPPAVGDTKS--GFGGI SVDAQNNVNVVATLN 346
QY 366 SWWPDITIFRSTGCGATWTRIDWMTSYNRSRLYVLDISABEPLWTFG-VQNPVPVSPKL 424
DB 347 RWFDEELIYRSTDAKTKWPIWDWNGYFNRTLYNLDYSAQPLDWGKTGVTFPDPLVKL 406
QY 425 GWMDEMAIDPFNSDRMLYGTGATLYATNLTFKWDSGGQIHIAPIVMYKLEETA VNDLIS 484
DB 407 GWMGDLLEIDPFNSDRMFPYGTATLYGTDLTNWDKGNVDISVKANGIEECANDVWVP 466
QY 485 PSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDYAEINPSIIVRAGSFDPS 544
DB 467 TKGAQLLSAVGDDCGFYHDDITKVPKMTTFNFSATTSIDYAESVPNFVRVGNVDTSK 526
QY 545 QPNDRHVAFTSDGGKNWFO-GSEPGVTTGGTVAASADGSRFVWAPGDPQGPVYAVGFG 603
DB 527 NQDKDCGISYDGGKNWFSAGNISGVYKAGTVAAGADAKTIVWSP-EGANNAAYTDNG 585
QY 604 NSWAASQGVPANAIQIRSDRVNPKTFYALSNGTFYRSTDGGVTF-QPVAAGLPSSGAVGM 662
DB 586 NKWTPCSGLPQGAQVSRSDRVNPKFYGLNGKFYISTDAGATFTQSSQTGLPKGK-GI- 643
QY 663 PHAVPGKEGDLWLA-ASSGLYHSTNGSSWSAITGVSSAVNVGFGKAPGSSYPVAFVVG 721
DB 644 PKTVIGHEDIIWAGDGLWHSTDSGATFTKVGSDVSDATVGLGKSKTDDGYPALYMDA 703
QY 722 TIGGVTGAYRSDDCGTTWVLINDDOHQYGNWQAITGDHANLRVYIGTNGRGIVYDGI 781
DB 704 TIDTAGIFRSDDEGATWVRINDDAHQYGSPPDYCITGDPNKYGRVFGVNGRGIVYGDID 763
QY 782 GAPSGSPSPSVS 793
DB 764 GS-QPTPTPSVT 774

RESULT 5
Q8TFP1 PRELIMINARY; PRT; 857 AA.
AC Q8TFP1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Endoglucanase C.
GN EGLC.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21914059; PubMed=11916668;
RA Haaper A.A., Dekkers E., van Mil M., van de Vondervoort P.J.I.,
RA de Graaff L.H.;
RT "EgIC, a New Endoglucanase from Aspergillus niger with Major Activity
RT towards Xyloglucan";
RL Appl. Environ. Microbiol. 68:1556-1560(2002).
DB EMBL; AY040839; AAK7727.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

```



DR InterPro; IPR000254; CBD\_fungal.  
DR InterPro; IPR002860; GH\_BNR.  
DR Pfam; PF02012; BNR; 7.  
DR Pfam; PF00734; CBM\_1; 1.  
DR ProDom; PD001821; CBD\_fungal; 1.  
DR SMART; SM00236; fCBD; 1.  
DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
SQ SEQUENCE 857 AA; 90459 MW; 482253ECD919C511 CRC64;

Query Match 35.7%; Score 1831.5; DB 3; Length 857;  
Best Local Similarity 44.4%; Pred. No. 7.2e-79;  
Matches 382; Conservative 126; Mismatches 306; Indels 47; Gaps 18;

QY 45 HAAATQPYTWSNVAI-GGGGFGVDFGVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDW 103  
DB 18 HAAASQAYTKWNVVTVGGGGFTGIVFNPSAKGVAVARTDIGGAYRLN-SDDTWTPMDW 76

QY 104 VGNWNGYGVVIAADPINTNKWAAVGMVYNSWDPNDGAILRSSDQATGATQITPLPFK 163  
DB 77 ANNSNHDWGDIAIATDPVDTDRVYVAVGMVYNDMPDNGSILRSTDQGTWETKLPFK 136

QY 164 LGGNMPGRCGERLAVDPNDNLLYFCAPSGKGLWSTDSGATWSQWTFPPDVTYIAP 223  
DB 137 VGNMFGRGVGERLAVDPNDNSILYFCARSGNGLWSTDYGETWSNVTAFTKWTGYFQDS 196

QY 224 TDTTGYQSDIQGVVVAFAFKSSSLGQASKTIIFGVADPNNPVFWSRDGATWQAVPGAP 283  
DB 197 SST--YTSDPVGIATWTFDSTSGSSGPTPRIIFGVVDTCESVFVEDAGETWTVSGEP 254

QY 284 T-QFIPHKGVDFPNVHLYIATSNTPGPDGSGDVWKFSTGVTGTRISPVSTDTAND 342  
DB 255 MYGFLPHKGILSPSEHTLYISYNGAGPYDGTGTGTVHKYNTSGVMTDISPTSMTDT-- 311

QY 343 YFCYSGLTIDRQHPNTIMWATQISWPDPTIIFSTDCGATWIRIWDTSYPNRSLRYLD 402  
DB 312 YCYGGLAVDLPQGTVMVAALNCWVDELIMWSTDSGGTWSPIWANGYPSINYYSYD 371

QY 403 ISAEPLWTFGVQNP--PVSPKLGWDEMAIDPFNSDRMLYGTGATLATNLTWKDWSG 461  
DB 372 ISNAPWLQDSTDFPV--RVGMVVEALAIIDPFSDHLYGTGTYGGHQLQWDS 428

QY 462 GQIHAPMKGLBEETAVNDLISPSPGAPLISALDGLGGFTHADVATVPSTFTSPVFTTG 521  
DB 429 HNVITSLAVGIEEMAVGLITPPGPAALLSAVGGDGGFVHTSLTTPAQYHYTPYSST 488

QY 522 TSVDYAEALNPSIIVRAGSDPPSQPNDRHVAFTDGGKWFQSGEPGVTGTGTAASAD 581  
DB 489 NGIDYAGNKPANIVRSGSD-----SDPTILASSFCESWYADYAASSSTATCQVALSAD 543

QY 582 GSRFVWAPGDPQPVVYVGVFGNSWAAQGVANAOIRSDRVNPKTFYALSNGTFRSTD 641  
DB 544 ADTILMNSDG-----AYRSANSATLVAVSLPSGAVIASDKANTTFYAGSGSFYLSDD 599

QY 642 GGVTFQVAAAGLPSSGAVGVMFHAPGKEGDLMLAAASSGLYHSTNGSGWSAI--TGVSSA 700  
DB 600 TAATP-TVTTTLGSSITANAI-RAQPSLAGDVWVSTDTGLFHSYNGKFTQIGSCCTEG 657

QY 701 VNVFGKSPAGSPAVFVVGITGGVTGAYRSDCGTWWLINDDQHQYGN-WGQAITGD 759  
DB 658 WSPFGKPSDGDYPLVFGFTVDTGLFKTEDQGVNMQIISDAEHGFGSASANVNGD 717

QY 760 HANLRVYLTNGRGVYDYGAPSGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSS 819  
DB 718 LQNYGKVFVGTNGRGVYD-----PSGT-LPSATATASSASTAVKSSSTSTSKVGSST 772

QY 820 PSSSP 869  
DB 773 TVSSSTATTITSSIKSTLTITTKSSSTSTSTSTSTSTSTSTSTSTSTSTSTST 831

QY 870 SGGVKQVQKN-----DSAPG 885  
DB 832 -SGWTCYENEVYSQCKSI 851

RESULT 6  
Q9P4T8 PRELIMINARY; PRT; 806 AA.

ID Q9P4T8  
AC Q9P4T8; 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE CEL6 protein precursor.  
GN CEL6.  
OS Agaricus bisporus (Common mushroom).  
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
OC Agaricales; Agaricaceae; Agaricus.  
OX NCBI\_TaxID=5341;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D649;  
RA Morales-Almora P., Thurston C.F.;  
RT "Molecular analysis of the cellulolytic genes in Agaricus bisporus";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ292929; CAC02964.1; --  
DR HSP; P00725; 2CBH  
DR GO; GO:000576; C:extracellular; IEA.  
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR00254; CBD\_fungal.  
DR InterPro; IPR002860; GH\_BNR.  
DR Pfam; PF02012; BNR; 8.  
DR Pfam; PF00734; CBM\_1; 1.  
DR SMART; SM00236; fCBD; 1.  
DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
KW Signal.  
FT SIGNAL. 1 23 POTENTIAL.  
FT CHAIN 24 806 CEL6.  
SQ SEQUENCE 806 AA; 84779 MW; 7C67382B62FF341 CRC64;

Query Match 34.7%; Score 1784; DB 3; Length 806;  
Best Local Similarity 43.7%; Pred. No. 1.2e-76;  
Matches 355; Conservative 131; Mismatches 277; Indels 50; Gaps 13;

QY 20 LAATAPFAVAALGVLPFIATSPAHAAATTPQYTSNVAI-GGGGFGVDFGVNEGAPGIL 78  
DB 1 MPSRVKFWITLTLTGTFSLV-----GQAASQSYRWQNVKIGGGGFGVFGIVENPSEKGLA 55

QY 79 VYRTDIGMYRWDANGRWIPLLDWGNWNGYGVVIAADPINTNKWAAVGMVYNSW 138  
DB 56 YARTDIGGAYKLN-ADDTWTELLDFADDSRWYGVDAIATDPVQPNRLYLATGYTNSW 114

QY 139 DPNDGAILRSSDQATGATWITPLPFKLGWMPRCGERLAVDPNDNLLYFCAPSGKGLW 198  
DB 115 DPNGHLLISSDYKTFDAAPLPFALGNMPCRCGERLWVDPNLSILYFCARSGNGLW 174

QY 199 RSTDGATWSQMTNFPDVGTYIANPTDTTGYQSDIQGVVWVAFDKSSSLGQASKTIIFV 258  
DB 175 KSTNSGRSWSKVSFTDTGSPVDPDSDTGLNSDKIGIAWVTIDKASGSGSATPRIFV 234

QY 259 VADP--NNPFWSRDGGATWQAVGAPGTFPHKGVDFPNVHLYIATNTGTPYDSSGD 317  
DB 235 VANKGNSIYISTNGGSSWSAVAGQPTSFPHKGVLSPSERALITYSDGGGPGYDGTGA 294

QY 318 VKFSPVSTGWTWTRISVPVSTDTANDYFCYSGLTIDRQHPNTIMWATQISWPDPTIIFRST 377  
DB 295 VYKISIDTGAWTDITPVGSDL-----PFGGGLADTQKNGVWVAALNSWDPGQIFRST 351

QY 378 DGGATWTRIDWTSYPNRSLRYLDISAEPLWTFGVQNPFPVSP--KLGMDSAMAI 434  
DB 352 NGCASWTPMDWAVYPTLNKYYSYNAALAPW-----IGNIVDVTPGNLQIGWMSLSID 407

QY 435 PFNSDRMLYGTGATLATNLTWKDSSGQIHIAPMVKLETAVNLDLSPSPGAPLISAL 494  
DB 408 PFDNSHWLYGTGETIYGRDLLKWDASHNVTIKSLADGVEETSVQALLISPPSPGPLVSAI 467

QY 495 GDLGGFTHADVATVPSTIFTSPVTGTSVDYAEALNPSIIVRAGSDFDSSQPNDRHAFS 554

```
Db 468 GDVGGFVHSLTRAPSAQFTNPXWTSADIDFAGNPNPNVRIIGTGSTS---GKQVAlS 524
QY 555 TDGCKNWFQSEPGVVTGCTVAASDGRFVWAPGDPQGVVYAVGFGNSAASQGVPA 614
Db 525 SDVGVTVNQHFGAPDNVQGGKVAISADADILLRTGNG----- 563
QY 615 NAQIRSDRVNPKTFYALNSGTFRYRSTDDGVTFOFVAAGLPSGAVGMFHAVPGKBDLW 674
Db 564 -VMVSRNQATFNIFYGASGKTFYVSTDMGKTFSAHGLSLGATSVDITVH--PSVSGDIW 620
QY 675 LAASSGLYHSTGSSWSAITGVSSAVNVCFKSAFGSSYPFAVVVCTIGVTVGAYSD 734
Db 621 ASTDKGLFHSDDGATFSALSGITQAWGVALGAPRSTGGYPFAVFAAANYGEVAYFRSD 680
QY 735 CGTWTWILNDHOYQ-NWQOATGDHANLRVYVIGTNGRIGVYGDIGGAPSGSPSPSVS 793
Db 681 RGVNWKINDAAGFGAASANCMAADPRVYGRVYIGTNGRIGYFGDVAGS---APPTSS 737
QY 794 PSASPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 826
Db 738 ATSTSVSTTTTTRSTTTTTRSTTTTTS-SGNGSP 769

RESULT 7
O74170 PRELIMINARY; PRT; 856 AA.
AC O74170;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Avicelase III.
GN AVIII.
OS Aspergillus aculeatus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurociates; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5053;
RN [1]
RP SEQUENCE FROM N.A.
RA Arai M., Takada G., Kawaguchi T., Sumitani J.;
RT "Avicelase III from Aspergillus aculeatus.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB015511; BAA29031.1; -.
DR PIR; T00349; T00349.
DR HSP; P00725; 2CBH.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR00254; C:BD fungal.
DR Pfam; PF02012; BNR; 7.
DR Pfam; PF00734; CBM_1; 1.
DR PRODOM; PD001821; C:BD fungal; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; C:BD FUNGAL; 1.
DR SEQUENCE 856 AA; 89620 MW; BE085983AF60ED76 CRC64;

Query Match 33.9%; Score 1740; DB 3; Length 856;
Best Local Similarity 43.5%; Pred. No. 1.5e-74;
Matches 366; Conservative 139; Mismatches 237; Indels 40; Gaps 18;

QY 30 AALGVLPALITASPAHAATTQPTWNSVNI-GGGGFVDGIVFNEGAPGILYVRTDIGMY 88
Db 4 SSLALLCAALGLKLAADAASQAYTWKVVGGGGGFTFPIVFNPSAKVAYARTDIGAY 63
QY 89 RWDAAANGRIPLLDVWG---WNWNGYGVVSIADPINTKNKVAAGVMTNSWDPNDGAI 145
Db 64 RLN-SDDTWPLMDVWGNDFTHDW--GIDALATDPVTDTRVYVAVGVMTNWDPNVGSI 119
QY 146 LRSDQGATWQITPLPKLGNMPCGRGRLAVDPNNDNLLYFGAPSGKGLWSTDSGA 205
Db 120 LRSTDQDGTWTETTKLPKFGVGNMPCGRGRLAVDPNKNLSILYFGARSGHGLWSTDYGA 179
```

```
QY 206 TWSQMTNPPDVGTIANPTDTTGYQSDIQGVVWVAFDKSSSSSLGOASKTI FVGADPNP 265
Db 180 TWSNVTSTFTWTTGTFQSSST--YTSDPVGAIAWTFDSTSGSSGSAITPRIFGVADAGKS 237
QY 266 VFVSRDGGATWQAVPGAPT-CFIPKGVFDPVNVHLYIATSNITGPGPYDSSGDVWKFST 324
Db 238 VFKSEDAGATWAWYSGBPQYGFPHKGVLSPEEKLIIYIYANGAGPYDGTNGTVHKYNI 297
QY 325 SGTWTRISPVSSTDTANDYFOYSGLTIDRQHPNTIMVATQISWPDPTTIFRSTDGATWT 384
Db 298 SGVWTDISP--TSLASTYCYGGLSVLDQVPGTILWAAALNCWNPDELIFRSTDGATWS 354
QY 385 RIWDTSTPNRSLYVLDISAEPLWTFGVQNP-PVSPKLGMDMAIDFNSDMLY 443
Db 355 PIWEMNGYPSINYYSYDISNAPITQDTSDDQFV--RVGMVYEAALIDFDSHMLY 411
QY 444 GTGATLATNDLTWKDSSGQTHIAPVMKGLBETA VNDLISPPSGAPLISALGDLGGFTHA 503
Db 412 GTGLTVYGGHDLTWDSKHNVTVKSALVIGIEMAVLGLITPPGFPALLSAGVDDGGFYHS 471
QY 504 DVTAVPSTIFTSPVFTTGSVDYAEINPSIIVRAGSFPSSQPNDRHVAFASTDGKWFQ 563
Db 472 DLDAAPNQAYHTPTVYGTGIDYAGKPSNIVRSASD--DYPT--LALSNFGSTWYA 526
QY 564 GSEFGVTTGTTVAASADGRFVWAPGDPQGVVYAVGFGNSAASQGVPAQAQIRSDRV 623
Db 527 DYAASTSTGTGAVALSADGDTVLLMSSTSGALVSKSQG--TLTAVSSLPAGVIAADKS 583
QY 624 NPKTFYALNSGTFRYRSTDDGVTFOFVAAGLPSGAVGMFHAVPGKBDLWLAASGLYH 683
Db 584 DNTVYGGAGAIYVSKNTAITSFTKTVS-LGSSTTVNAI-RAHPSIAGDVWASTDKGLWH 641
QY 684 STNGGSSMSAI-TGVSSAVNVGFGKSAFGSSYPFAVVVCTIGVTCAYRSDDCGTTWLI 742
Db 642 STDYGSTFTQIGSGVTAGWSFGKASSTGSVVIYVGPFTIDGAAGLFRKSDAGTNWQVI 701
QY 743 NDDQHOYQNWQOA--ITGDHANLRVYIGTNGRIGVYGDIGGAPSG-----SPSPSV 792
Db 702 SDASHGFGS-GSANVWNGDLQTYGRVFRGHERPGHLLRQSRPAGRHGDGDDGDTTSTKT 760
QY 793 SPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 852
Db 761 STTVSTTLTKTSSASTSSSTTVKTTTSSSTTSKASSTTTTKTTTSSSGTATYA 820
QY 853 SS 854
Db 821 SA 822

RESULT 8
Q729M8 PRELIMINARY; PRT; 838 AA.
AC Q729M8;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Cel74a.
GN CEL74a.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaeae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-QM6a;
RA Foreman P.K., Brown D., Dankmeyer L., Dean R., Diener S.,
RA Dunn-Coleman N.S., Goedegebuur F., Houfek T.D., England G.J.,
RA Kelley A.S., Meerman H.J., Mitchell T., Mitchinson C., Olivares H.A.,
RA Teunissen P.J.M., Yao J., Ward M.;
RT "Transcriptional Regulation of Biomass-Degrading Enzymes in the
RT Filamentous Fungus Trichoderma reesei.";
RL J. Biol. Chem. 0:0-0(2003).
DR EMBL; AY281371; AAP57752.1; -.
```

SQL SEQUENCE 838 AA; 87132 MW; FBE3D08F2A74FA84 CRC64;  
Query Match 31.7%; Score 1625.5; DB 3; Length 838;  
Best Local Similarity 41.7%; Pred. No. 3.9e-69;  
Matches 350; Conservative 131; Mismatches 296; Indels 63; Gaps 23;  
QY 25 SPVAALGVLPATITASPAAHATTPQYTSNVAI-GGGFGVGVGVNEGAPGLIYRTD 83  
DB 4 SRVALVLGAV-----IPAHAA-----FSWKNVKGGLGGGFGVPGIIPHPKTKGVAYARTD 53  
QY 84 ICGWYWDANGWIPILDHW-----GWNNGYGVSVIAADPTNTNKNVAAVGYTNSWD 139  
DB 54 IGGYRLN-ADDSWTAVTGDIANAGHWN---GIDVALDPQDDQKQYAAVGYTNSWD 109  
QY 140 PNDGAILRSDGATGATGATITPLPKLGNMPCRGMRGERLAVDPNNNDNLYFCATSGKGLMR 199  
DB 110 PSNGAILRSDGATGATGATGATITPLPKVGNMPCRGMRGERLAVDPNSNLIYFCATSGKGLMR 169  
QY 200 STDGATWQMTNFPDVGTYIANPTDTTGYOSDIOGVVWVAFKSSSLGQASKTIFVGV 259  
DB 170 STDGVTFSKVSFTATGTYIPDSGNSYNSDKQLMWVTFDSTSTTGATSRIFVGT 229  
QY 260 ADP-NNPVFWRSRGGATQAVPGATGFIPIHKGVDVFNHVLVATNTGPGVDGSSGDV 318  
DB 230 ADNITASVYVSTNAGSTWASVPGQPKYFPHKAKLOPAEKALYLYTSDGTGPGYDGLTGSV 289  
QY 319 WKFSVTSGTWTRISVPVSTDTANDYFCYGLTIDRQHPNTIMVATQISWMPDTIIFRSTD 378  
DB 290 WRDYDAGTWDITPVSGSL---YFGGLGDLQPKGLVVASLNSWMPDQALFRSTD 346  
QY 379 GGATWTRIMDWTSPYRSLRYVLDISAEPWLTFG-VQPNPVPSPS---PKLGMWDEAMAI 433  
DB 347 SGITWSPWAWASYPTEYYISITPKAPWIKNNFIDVTSESFSDGLIKELGMWIESLEI 406  
QY 434 DPNSRMLYGTGATLYATNDLTKWDSGGQIHAPMVKGLBETAVNDLISPPSGAPLISA 493  
DB 407 DPTDSNHWLYGTGWTIFGGHDLTNWDRHNVISQSLADGTEEFVQDLASAPGSELAA 466  
QY 494 LGDLGGFTHA---DVTAVPSTIFTSVPVFTGTSVDYAEINPSIIVRAGSFDPPSQPNDRH 550  
DB 467 VGDNDGFTFASRNDLGTSPQVWATPTWATSTVDYAGNSVKSVVRVGTAGTQQ-----521  
QY 551 VAFSTDGKKNWFOGSPBPGVTTGGTVAASADGSRFVWAPDGPQPVVYVFGFNGSMAASQ 610  
DB 522 VALSSDGGATSIDYAADTSMNGGTWAYSDGDTILWSTASSG---VQRSQFOGSPASVS 578  
QY 611 GVPANAQISDRVNPKTFFVALSNGTYRSTGDTGTFQPVNAAGLPSSGAVGM--FHAVPG 668  
DB 579 SLPAGAVIASDKKTNVFFVAGSGSTFVSKDTGSSF---TRG-PKLGSGTIRIDIAAHT 634  
QY 669 KEGDLMAASSGLYHSTNGSSWSAI-TGVSSAVNVGFGKSAAGSSYPVAVVGTIGGVT 727  
DB 635 TAGTLVSTVDGIFRSTDSGTTFGQVSTALTNTYQIALGVGS-GSNW-NLYAFGT--GPS 690  
QY 728 GA---YRSDDCGTWVLINDDHOQYGNWGOAITGDHANLRVYIGTNGRGVY--GDIGGA 783  
DB 691 GARLYASGSGASWTDIQQSQGFGSDSTKVASGSGTAGVYVGTNGRGVYAGTIVGG 750  
QY 784 PSQSPSPSPSPASPSLSP 843  
DB 751 TGGT-----SSSTKQSSSTSSASSTTLRSSVSTTRASTVTSSTSSAAGTGS 801  
RESULT 9  
Q82K30 ID PRELIMINARY; PRT; 739 AA.  
AC Q82K30  
DT 01-JUN-2003 (TEMBLrel. 24, Created)  
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)  
DE 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
DE Putative glycosyl hydrolase.  
GN SAV2574.  
OS Streptomyces avermitilis.

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
OC NCBI TaxID=33903;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=21477403; PubMed=11572948;  
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,  
Kikuchi H., Shiba T., Sakaki Y., Hattori M.;  
RA "Genome sequence of an industrial microorganism Streptomyces  
RT avermitilis: deducing the ability of producing secondary  
metabolites";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN-MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=22609306; PubMed=12692562;  
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
Sakaki Y., Hattori M., Omura S.;  
RA "Complete genome sequence and comparative analysis of the industrial  
RT microorganism Streptomyces avermitilis";  
RL Nat. Biotechnol. 21:526-531 (2003).  
DR EMBL: AP005031; BAC70285.1; -;  
GO: GO:0016787; F:hydrolase activity; IEA.  
DR InterPro: IPR002860; GH\_BNR.  
DR Pfam: PF02012; BNR; 9.  
DR TIGRFAMs: TIGR01409; TAT\_signal\_seq; 1.  
KW Hydrolase; Complete proteome.  
SQ SEQUENCE 739 AA; 78399 MW; 8E3B0F3D0F5B3D28 CRC64;  
Query Match 28.1%; Score 1442; DB 16; Length 739;  
Best Local Similarity 40.8%; Pred. No. 1.6e-60;  
Matches 328; Conservative 107; Mismatches 272; Indels 96; Gaps 25;  
QY 7 IRLTMSRRLLSLAATASFAVAALGVLPATITASPAAHATTPQYTSNVAIAGGGG 63  
DB 1 MRPSPSRRTV--LAGTAA---AAALTAVP-AVGSQAHAAETTAGPSYRWNVIGGTG 53  
QY 64 FVDGIVNEGAPGLIYRTDITGGMYRDAANGRWIPLLDVGNWNNWNGVYVSIADPN 123  
DB 54 FVTGVLFHPSVRGLAYARTDIGAYRWDGARGWTPLIDHLGWDNDNLLGVAMAVDPHT 113  
QY 124 TNKWAAGWYTNWDNDGAILRSSDQATGATITPLPKLGNMPCRGMRGERLAVDPNN 183  
DB 114 PDRLYLAVGTVAQSWAGN-GAVLRSEDRGATWTRTDLTVKLGGNEDRGAGERLLVDPRD 172  
QY 184 DNILYFGAPSGKGLWRSTDSGATWQMTNFPDVGTYIANPTDTTGYSDIQGVVWVAFDK 243  
DB 173 SDTLWLGT-RHDGLLKSTDRGATWAAATAP-----AKANSSGQGVVF-----214  
QY 244 SSSSLGQASKTIFGVADPNP-----VWFRDGGATWQAVGAPTGF---IPHKGVFDP 295  
DB 215 ----LVAAGRIVYAGWGDGDTSGTANLYRTAD-GTTWGAVPGRPSGTSAKVLRAAYDT 269  
QY 296 VNHVLYATNTGPGYDGSQDVKFVSTGTSWTRISPV-----PSTDANDYFCYGLTI 351  
DB 270 HTRLEYTYGDAPGGQSDGSHVHLKRTATGTATGTEVTPVKPGGTSDGSADTFAYGGVAV 329  
QY 352 DRQHPNTIMVATQISWMPDTIIFRSTGCGATWTRIDWMTSYNRSRLRYVLDISAEPWLT 411  
DB 330 DARPGTLVSTNNRWADGDTVFRSTDGRTWTSKLD-----AAVFDVSETFFLDW 380  
QY 412 GVQPNPVPSPKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHAPMVK 471  
DB 381 GDD-----KPKFCWMIQALAVDPVDSQHVYVGTGATLYGTRDLKRW-----APRIR 426  
QY 472 GLEETAVNDLISPPSG-APLISALGDLGGFTHADVAVPST-IFTSVPVFTTGTSDVYABL 529  
DB 427 GLEESAVRQLISPPVGEAHLISGLDIGWYHERLTASPSRGMATNPVFGSAGTAAQAAA 486  
QY 530 NPSTIIVRAGSFDPPSQPNDRHVAFPSTDGGKWN--FQSGSEPGVTTGGTVAASADGSRFW 587

```

147 RSSDQATQITLPLPKLGGNMPGCMGERLAVDPNNDNIIYFGAPSGKGLWRSTDGAT 206
141 RSFDRGQFERADUPFKLGGNQLGRANGERLAVDPHDGRVLLLSGRDA-GLWRSDRGAAH 199
207 WSONTFNPDVGTIYIANPTDTTGYQSDIQGVVWVAFDKSSSSISLGQASKTIFFGVADPNPNV 266
200 WARVEGFPADALACATARNHVGSQAV-GIAFVVFDAASGHAGATPRIYVGVSTAQTSI 258
267 FWSRDGATQWAVPGATGFIPIHKGVPDPVNHV-----YIATSNCTGGPYDGSQGVW 319
259 YVSEDAKRTWSAVAGQPKGLRP-----SHWVGHNAQWYLSYGDPRPGDLMAGGALW 310
320 KFSVTSCTWTRISVPSTDTFANDYEGYSLGIDRQHPNTIMVATQISWHPDIIIRSTDG 379
311 NYDATQGRWREISPIQPAT-GDGFVGAVAVDPQHPQVLLASTPRRRTPRDEVFRSGDG 369
380 GATWRIWDWTSYPNRSRLRYVLDISABPWLTFGVQPNPVPSPKLGWMDMAAIDPFNSD 439
370 GRSWVPLAAAOQF-----DHSAAFW-TAHATPH-----WIG-ALAIIDPFDSN 409
440 RMLYGTGATLYATNDLTWKDGGQIHIAPMVKGLEETA VNDLISPSPGAPLISALDGLGG 499
410 HATFVTGVIWASRNLTQFSEQQPLQWFWQDRGLEETVPLDLSPMAGAHLLSALGDIDG 469
500 PTHADVTAVPSTIFTSTPVTGTSVDYAEIAPSIIVRAGSFPDPSSQPNDRHVAFSTDGK 559
470 FRH-DALDTAQLQYLGRLTNGESIDGAGQAPQWVYRSQTV-RDRNNEIRALYSQDGT 527
560 NWFQ-GSEPPGGVTTGTTVAASADGSRFVWAPGDPQPVVYAVG-FGNSWAASQGVPAQA 617
528 HWAAPASEPPRGQAGTITAIADASQVWVPDQGG---VMRTGDFGKRWQVQGLPDTAV 584
618 IRSRVNPKTFYALS--NGTFYRSTDDGVTFOFVA--AGLPSSGAVGVMFHVPKKEGDL 673
585 VVADRVDAQRWYAADRVSGRLYESDGAASPRDTGQGVSPARDE-----RARPQLRDP 639
674 WLAASSGLYHSTNGSSWS-----AITGVSSANVGVFGKSPAGSSYPVAVFVGTGTGVTG 728
640 WRAGVVLASPTLGVMRWQQLRTLKSKSDEARSIGIKALRAGAPPALYLAGRVAGVDG 699
729 AYRSDDCGTTWVLINDDOHOGNWOQATIGHANLRRVYIGTNGRGIVYGD 779
700 IFRSDDEGAHRRINDDAHFRGK-PYSVTGDPRIAGRVTFATGGRGIFYGD 749

RESULT 11
Q9WYE1 PRELIMINARY; PRT; 707 AA.
ID Q9WYE1 AC Q9WYE1
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Endoglucanase, putative.
GN TM0305.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eissen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL; AE001712; AAD35393.1; -.
DR F1R; F72393; F72393.
DR TIGR; TM0305; -.

```

```

487 REAYVVRGTGWDHNG-----AYSHDGGRTWAPFEAQDPIADKAPGPIATSDAGGTLW 540
588 A-----PGDPCQPVVYA-----VGFENWAAASQGVPAQAIRSDRVNPKTFYA--LSNGTFY 637
541 SFVHWDG-----TTAAHRSTNDGASWSESVSPFKGATPVADPFRFVAYDFDNGTLY 595
638 RSTDGVTQFVVAAGLPSSGAVGMFHAVPGKGDWLAAAS-SGLYHSTNGGSSWSAITG 696
596 ASTDSGRSFARAGGLP-SGDSQFKLVAAPGRSGDLWLSAKMNGLYRSTDGGDTFARIDS 654
697 VSSAVNVGSKSAPGSSYPVAVFVGTGIGVYRSDDCGTTWVLINDDOHOGNWOQAI 756
655 CWASYTLGFGKAADGADYPAIYQVSTETTIVYRSDDAARTWVRINDDAHQWNGICEAV 714
757 TGDHANLRRVYIGTNGRGIVYGD 779
715 VGDPRHGRVYIATNGRGIOYGE 737

RESULT 10
Q8PLM5 PRELIMINARY; PRT; 751 AA.
ID Q8PLM5 AC Q8PLM5
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Cellulase.
GN CELA OR XACI770.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chamberg J., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White P.P.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
RL EMBL; AE011809; AAM36634.1; -.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 7.
KW Complete proteome.
SQ SEQUENCE 751 AA; 81259 MW; D4134D96285E599D CRC64;
Query Match 22.9%; Score 1173.5; DB 16; Length 751;
Best Local Similarity 35.5%; Pred. No. 8.1e-48;
Matches 274; Conservative 117; Mismatches 319; Indels 61; Gaps 20;

```

```

27 AVAAALGVLPITATSPAHAAATQPTWTSNVAIGGGGFGDGVFNEGAPGILYVRTDIGG 86
22 AVCLTLLVFWTSEASDEPTDPPQWHSVAIGGGGFGVTVGLFHPAERDLAYARTDVG 81
87 MYRWDAANGRWIPLLDVGNWNGYGVVSIADPINTNKWAAVGYTNSWDNDGAIL 146
82 AYRWDAQQWQVALTDWLGADDDNLMGIDAFVADPADPNALYLAAGTYMHERAGN-AVL 140

```

```
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 9.
KW Complete proteome.
SQ SEQUENCE 707 AA; 79496 MW; 8CD8743CBDAGA99F CRC64;

Query Match      22.6%; Score 1163; DB 16; Length 707;
Best Local Similarity 35.2%; Pred. No. 3e-47;
Matches 264; Conservative 123; Mismatches 275; Indels 89; Gaps 26;

Qy 52 YTSNVAIGGGFVDGIVFNEGAPGILYVTRDIGMYRWDAAAGRWIPLLDWMGNWNGY 111
Db 21 FEKSVIEINGGFFVGLIIFPASPGLYATDVGLRWBEETKRWQLDFLRDQSDY 80
Qy 112 NGVVSIAADPINTKNKWAAGMTNSWDNDGAILRSSDQATWQITPLP-----FKLGNM 168
Db 81 MGVLVALDPSDPKRIYAMTKYTQDW-AGYGAILISEDYGETWTIVLNDKYKIGKVGNE 139
Qy 169 PGRGMRGLAVDNDNNILYFGAPSGKGLWRSTDSGATWQMTNPPDVGYTIANPTDTTG 228
Db 140 DGRNAGERLQVDNFSSVLFMGT--TKYGLWKSDFGKNWKKVDSFPST-----186
Qy 229 YQSDIQGVVAVFADPKSSSIQSQAKTIFVGADPNPNVFWSRDGGATWQAVPGAPTGP 288
Db 187 -----SVTFVLFDEKSGEKGSPTRIFVGCSEPKG-IFVTEDGGTTWNVLNPLNDLIP 239
Qy 289 HKGVFDVNVHLVLIATSNCGPYDGGSDVWVKFSVTSCTWTRISVPSTDTANDYFGYSG 348
Db 240 LRKGIH--DGLIVLISNALGPNGATGAVWKYVIADQKVDVTPMKGD-----FGYCG 291
Qy 349 LTIDRQHPNTIMVATQISWHPDITIFRSTDGGATWTRIMWTSYVNRSLRYLDISAEPW 408
Db 292 IDVQE--NVVISTLDWYPHDEIFISLNGGETWRPLLEKANF-----DINKAPW 339
Qy 409 LTFEQVQNPVPVFPKGLWMDMAIDPFPNSDRMLYGTGATLYATNLTWKWDSGGQIHAP 468
Db 340 IK--DILNP-----HWISD-VKIDPDMNRAIFTGCGVWVYELKSFEG----MGK 384
Qy 469 MVK-----GLETAVNDLISPPSGAPLISALGDLGGFTHADVAVPSTTFTSPVFTGT 522
Db 385 PVKWFENRGLTEETVVLQLPPIGERPLLSAIDAWGGRHSLDTPPSSMT-KPLKWTSL 443
Qy 523 SDVYAEINLPSIIVRAGSFDPSSQPNDRHVAFTDGGKWNFQ-GSEPGGVTTGG--TVAA 579
Db 444 GIAFAYQNSKFVARVHTY---TYP--FLSYSEDGGINWEIETVPSGITDGGRLSLAVS 497
Qy 580 ADSRFFWAGDGPQVPVAVGFGNSWAASCGVPA---NAQIRSDRVNPKTFVAL--SNG 634
Db 498 NDGKTLWSPAN--HEVIVSSDKGSKWKAISVPVPEFNYFPASDPVNPSPKFYFDWKNG 555
Qy 635 TFYRSTDDGVTTFQPVAGLPSSGAVGMFHA---VPGKGDWLAAAS-SGLYHSTNGGS 689
Db 556 DFLISKDGGKSFMK-GAKLPFNNWVSLYFPVPLADPRDGIWIALQWNLGRSKDGGI 614
Qy 690 SWSAIGTVSSAVNVRGKSPAGSSYPVAVFVVGITGGYATRSDDCGTWTWLINDQHQY 749
Db 615 TFEELGNVDIAYVIGAPKPGTDYPAIYLNGVWNGYVIGIFMSDTECKTWIRINNDKHQF 674
Qy 750 GNGGQAITGDHANLRRVYITNGRGIVYGGI 780
Db 675 G-WHYVMIGDMNBEGRIFLGTGEGRIIVGSEV 704

RESULT 12
Q8P9U5 ID Q8P9U5 PRELIMINARY; PRT; 751 AA.
AC Q8P9U5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cellulase.
GN XCC1752.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

Xanthomonadaceae; Xanthomonas.
NCBI_TaxID=340;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
EX MEDLINE=2202145; PubMed=12024217;
da Silva A.C.R., Ferro J.A., Reinach P.C., Farah C.S., Furlan L.R.,
Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
Camarotte G., Cannavan P., Cardozo J., Chambergo F., Ciapina L.P.,
Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
Lacali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
Trindade dos Santos M., Truffi D., Teai S.M., White F.P.,
Setubal J.C., Kitajima J.P.;
"Comparison of the genomes of two Xanthomonas pathogens with differing
host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AB012276; AAM41043.1; -.
GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
GO; GO:0006457; P:protein folding; IEA.
GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001179; FKBP_PPIase.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 7.
DR PROSITE; PS00453; FKBP_PPIASE 1; 1.
DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
KW Complete proteome.
SQ SEQUENCE 751 AA; 81140 MW; 7378B6AA0B5698A4 CRC64;

Query Match      22.4%; Score 1151.5; DB 16; Length 751;
Best Local Similarity 34.6%; Pred. No. 8.9e-47;
Matches 273; Conservative 127; Mismatches 305; Indels 83; Gaps 23;

Qy 23 TASFAVAAALGVLPITATASP-----AHAATTQPTWNSVAIGGGFVDGIVFNEGAPGIL 78
Db 14 TTTGAWLAVLSL-LLLFSTPNVRAAEPATSGPYQWRSVAIGGGFVTGLVFPHERGLA 72
Qy 79 YVRTDIGMYRWDAAAGRWIPLLDWMGNWNGYVSIADPINTKNKWAAGMTNSW 138
Db 73 YARTDVGAYRWDAQAQQWTALTDLWGADDWNLMDIDAFADVPADADALYLAAGTYMHER 132
Qy 139 DPNDGAILRSSDQATWQITPLPFLKGNMPPGRGMRGLAVDPNNDNILYFGAPSGKGLW 198
Db 133 AGN-AAVLRSFNRRGTFERADLPFLKGNQQLGRANGERLAVDPHGRVLLLSRDA-GLW 190
Qy 199 RSTDSGATWSQMTNPPDVGYTIANPTDTTGYQSDIQGVVWVAFDSSSSSLGQASKTIFVG 258
Db 191 RSDRGAAHWAKVAFSPDAAALAGATARNHVGREQAV-GIAFVVFDAASGNTGTPTRIYVG 249
Qy 259 VADPNPNVFWSRDGGATWQAVPGAPTGPFPKHGVPDPVNVHLYIATNSNTGGPYDGGSDV 318
Db 250 VSTEQTSLYVSEDAGRSWAPVAGQPRGLRPSHMGSGDGH-WYLSYGDQPGPDLMAAGGAL 308
Qy 319 WKFSVTSQWTRISVPSTDTANDYFGYSGLTIDRQHPNTIMVATQISWHPDITIFRST 378
Db 309 WKFTPAQGRWREISPIQO-PASGDGFGWGAVDVDPQVQLASTFRRTTFRDELRSVD 367
Qy 379 GGATWTRIDWTSYVNRSLRYVLDISAEPLWITFGVQPNPPVSPKLGWMDMAIDPFNS 438
Db 368 GKKWAPL-----LADAVFDSRAPW-TAHATPH-----WNG-ALAIIDPDG 407
Qy 439 DRMLYGTGATLYATNLTWKWDSGGQIHAP-----MVKGLETAVNDLISPPSGAPL 490
Db 408 NHALFVTGYGIWASRNL-----QDFAAPQRLQWFWQDRGLEETVPLDLLSPMAGAH 460
```

```
QY 491 ISALGDLGGFTADVTAVPSTTFTSPTVTTGTSVDYAEALNPSIIVRAGSPDPSSQPNDRH 550
Db 461 LSGAGLIDGIDRDELDR-OLQVAGPRLTNGESIDAGQAPQWVRSGTV-RDRRNEIR 518
QY 551 VAFSTDDGKNWFQ-GSEPGVTTGGTVAASADGRFVWAPDGPQPVVYAVGFGNSWAAS 609
Db 519 ALYSRDRGGKQWAFAPSEPPAGQAGSITAIGADAQVVMAPERGN--WRTSDFGAQQRV 576
QY 610 QGVPAQAIRSDRVNPKTFYA--LSNGTFYRSTGGVTFQFVAAGL----- 653
Db 577 DGLPDTAVVMAVDARRWADVASQLYESTDAARSFR--ATGVQVGSPPARDERTRPQ 634
QY 654 --PSSGAVGMFHAVPGKEGDLMLAASSGLYHSTNGSSWSAITGVSSAVNNGFGKAPG 711
Db 635 LRDPWRAGVVYLASPGK-----GVMRQDG--TLQVLSQPDPEARSLGTGKALRA 682
QY 712 SSYPAPVVTGTTGGVTCAYESDDGTTWVLINDDOHQYGNWGOAITGDHANLRVYIGTN 771
Db 683 GAPPALYLAGRVGVGDGFRSDDGQVQWRINDDAHRFR-PYSVTGDPRIAGRVPYATG 741
QY 772 GRGIVYGD 779
Db 742 GRGIFYGD 749

RESULT 13
Q8J0D2 PRELIMINARY; PRT; 812 AA.
ID Q8J0D2 PRELIMINARY; PRT; 1779 AA.
AC Q8J0D2;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Oligoxylucan reducing end-specific cellobiohydrolase.
OS Geotrichum sp. M128.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; mitosporic Dipodascaceae;
OC Geotrichum.
OX NCBI_TaxID=203496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M128;
RX MEDLINE=22359052; PubMed=12374797;
RA Yaci K., Mitsuishi Y.;
RT "Purification, Characterization, Cloning, and Expression of a Novel
RT Xyloglucan-specific Glycosidase, Oligoxylucan Reducing End-specific
RT Cellobiohydrolase.";
RL J. Biol. Chem. 277:48276-48281(2002).
DR EMBL; AB089343; BAC22065.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 9.
KW Hydrolase.
SQ SEQUENCE 812 AA; 87080 MW; F5C740F32D8B17B5 CRC64;

Query Match 21.9%; Score 1125; DB 3; Length 812;
Best Local Similarity 34.5%; Pred. No. 1.7e-45;
Matches 285; Conservative 129; Mismatches 291; Indels 120; Gaps 31;

QY 31 ALGVLPATASPAHAATTPQYWSNVAIGGGVFDGIVNEGAPGILYRTDIGNYRW 90
Db 6 SLKALTASILASLAVAKEHYEFKNAIGGGVITGIVAHPTKOLLVARTDIGNYRW 65
QY 91 DAANGRWIPLLDVGNWNGVGVSTAADPINTKVAAGVMT-NSWDPNDGAILRSS 149
Db 66 DACTSKWIPLNDTIEADNMINGTESIALDPNPNDRLLYLAQGRVYGDW-----AAFYVSE 121
QY 150 DQAGTWOITPLPKLAGNMPGRGGERLAVDPNNNDILYFCAPSGKGLWRSTDGATWSQ 209
Db 122 DRQGSTIYESPPMGANDGRNGERLANVNPNSNEVWGTRT-EGIKSSDRAKTWTN 180
QY 210 MTNFPDPVGTIANPTDTTGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNFWS 269
Db 181 VTSIPDAF-----TNGIGYTS-----VIFDP-----ERNGTIYASATAPOG-MVYT 220
```

```
QY 270 RDGATWQAVPGAPTGF-----PHKGVDPDPNVHVIYIATSNITGGPY 311
Db 221 HDGVSWEVPAGQPSWLNRTTGAPDPKPAIAPQPMKVALTP--NFLYVTVADYVPGW 278
QY 312 DGSSEDVWKSIVTSGTWTRI-----SPVSTDTANDYFGYSGLTIDRQHPNTIMVATQ 364
Db 279 GVTGFEVWRQNRSTGAWDDITPRVGNSSPAPYNNOTFPAGFCGLSDVATNPNRLVVIIT- 337
QY 365 ISWMPDTI---IFRSTDCGATWTRIWDWTS-----YFNRSLRYVLDISAEPWLTFG 412
Db 338 LDRDPGALDSIYLTSDAGATWKDVTLQSSPSNLEGNHGHPTNAARY-KDGPVVPWLDEN 396
QY 413 VQP-----NPVPSP---KLGMDEAMAIIDPNSDRMLYGTGATLYATNDLTK-----WDS 460
Db 397 NGPQWGGYGAPHGTPGLTKFGWMSAVLIDPPNPEHLMYGTGATWATDLSRVEKDW-- 454
QY 461 GGQIHIAP---MWKGLEETA VNDLI SPSPGAPLISALGDLGFGFTHADVTAVPSTIFTSP 516
Db 455 -----APSWYLQIDGIEENAILSLRSPKSGAALLSGIGDISGMRHDDLTK-PQKMFAP 507
QY 517 VFTTGSTVDYAEALNPSIIVRAGSPDPSSQPNDRHVAFTSDGGKNW--FGSPPGGVTT-- 572
Db 508 QFSNLDSDIAAGNPNVVRAGSSGHEYDSACARGAYATDGGDAWTFPTCPGEMNASHY 567
QY 573 -GGTVAASADGSRFVWAP--GDPGQPVVYAVGFGNSWAASQGVPA-----NAQIRSDRV 623
Db 568 QGSTIAVDASGSIQVWSTKLDEQASGPMWYSHDYGKTSW---VPAGDLKAQATANVLSDKV 623
QY 624 NPKTFYALNSGTFYRSTGGVTFQFVAAGLPSGAGVGMFHAVPGKEGDLWL-AASSGLY 682
Db 624 QDGTFYATDGGKFFVSTDDGGSYAAKAGLVLT--QTSLEMPAVNPWVAGDVMVPEGGGLF 681
QY 683 HSTWGGSSWSAI-TCVSAVNVGFKS---APGSSYPAPVFFVGT--IGVTCAYRSDDC 735
Db 682 HSTDFGASFTRVGTANATLVSVGAPKSDGKASAPSAVFITGTDKPSDGLYRSDDN 741
QY 736 GTTWVLINDDOHQYGNWGOAITGDHANLRVYIGTNGRIGVYGD 780
Db 742 GSTWTRVNDQEHYSG-PTMIEADPKVYGRVYLGTNGRIGVYADL 785

RESULT 14
O52374 PRELIMINARY; PRT; 1779 AA.
ID O52374;
AC O52374;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Family 10 xylanase (EC 3.2.1.8).
GN XYN.
OS Caldicellulosiruptor sp. Rt69B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=70295;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rt69B.1.
RA Morris D.B., Gibbs M.D., Ford M., Thomas J., Bergquist P.L.;
RT "Family 10 and 11 xylanase genes from Caldicellulosiruptor sp.
RT Rt69B.1.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; AF036924; AAB95326.1; -.
DR FIR; T31085; T31085.
DR HSSP; Q06851; INBC.
DR GO; GO:0030246; F:carbohydrate binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; F:carbohydrate metabolism; IEA.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR00584; CBD IV.
DR InterPro; IPR005084; CBM_6.
```



DR InterPro; IPR003305; CBM\_Cenc.  
DR InterPro; IPR008965; Cellul\_bind.  
DR InterPro; IPR008979; Gal\_bind\_like.  
DR InterPro; IPR001000; Glyco\_hydro\_10.  
DR Pfam; PF00942; CBM\_3; 3.  
DR Pfam; PF02018; CBM\_4; 9; 2.  
DR Pfam; PF03422; CBM\_6; 1.  
DR Pfam; PF00331; Glyco\_hydro\_10; 1.  
DR PRINTS; PD00134; GLYDRLASE10.  
DR ProDom; PD001947; CBD\_3; 3.  
DR SMART; SM00606; CBD\_IV; 1.  
DR SMART; SM00633; Glyco\_10; 1.  
DR PROSITE; PS00591; GLYCOSYL\_HYDROL\_F10; 1.  
KW Glycosidase; Hydrolase; Xylan degradation.  
SQ SEQUENCE 1779 AA; 194304 MW; CE5269B6806B5CED CRC64;

Query Match 8.4%; Score 431.5; DB 2; Length 1779;  
Best Local Similarity 19.8%; Pred. No. 2.7e-12;  
Matches 252; Conservative 162; Mismatches 382; Indels 477; Gaps 53;

QY 65 VDGIVFNEGAPGILYV----RTDIGMYRWDANGRIPIILLDWMNNWNGYGVVISAAD 120  
DB 20 VGTILFHOEAKAAAYVDPECTDTLSFAVCKSNIA-----VDMGN-AYNGKSSIRVS 71  
QY 121 PINTNKVMAAVGMVTSNDPNDGAILRSSDQATWQIT-----PLPFKL-----G 165  
DB 72 --NRSSIDWGAV-----DVKNIMNGTTWVVSAYVKHSYQKPVAFGISAVYDDG 119  
QY 166 GNMGRGMRGLAVDPNDNILYFGAPSGKGLWRSTDSGATWSQNTWPPDVGTVIAPTD 225  
DB 120 SGVKSTLIGEVVAI-PN-----YWKXI-----VGRKTPNISN 150  
QY 226 TTGQSDIQGVVWVAFDKSSSLGQAKTIFGVADPN---NPVFWSR---DGGAT--WQ 277  
DB 151 VRNLLIVVHTIVSGVDYND-----YIQMDNYSLSNAVTFSSGFESGTEGQW 201  
QY 278 A----VPGAPTGPTRHKGVPDPVNVHLYIA--TSNTGG---PYDG---SSGDVWKFVS--- 323  
DB 202 ARGSGVTVPKDSVAVNGKYS-----LYSGRTSNWGAQIPVDITILEQGVKYSVWVY 256  
QY 324 -TSGTWTTRISVPSTDPAND-YFGYGLGTLDRQHPNTIMVATQISWMPDTIIFRSTGGA 381  
DB 257 QNSGSTQKMSLTWRRFATDPSTSYENLYNDRVPSNTWVPEPSGY-----SIPAGV 308  
QY 382 TWTIRWDWTSYPNRSRY-----VLDIS--AEP----- 407  
DB 309 TVSELLEYBEAQANLAFWVDLKIYDLSKLAPEWEIPSLIEKYDYFKYGVALSYSKI 368  
QY 408 -----WLTFGVQPNPPVPSPKLGW 426  
DB 369 ASDTEKRWLKHFNSTAGNEMKPSSELLISENNYNFKADEFVNFAFNSNIAIRGHTLVW 428  
QY 427 MDEA-----MAIDP----- 435  
DB 429 HEQTPDWFKANGNTLSKDALSLRKLQYIVTVVGRYKGVYAVDWVNEAIDESQGNFR 488  
QY 436 -----FNSDRMLYGTGATLYA-----TNDLTQWDSGGQIHAPWVKGLEETAVNDL 481  
DB 489 RSNWYNICGPEYIEKAFIWAHEADPDADKLFYNDYNTENSQKRFYNNIKSLKE----- 542  
QY 482 ISPPSGAPLISALDGLGGFTHADVTAPSTIFTSVPVFTT--GTSVDVVAELNPSIIVRAGS 539  
DB 543 ----KGVF-IHGIG-LOCHINLWPSISENTIKLFSSIPGLEIHTELDMSFYQWGSS 596  
QY 540 FDPSSQPN-----RHVAF-----STDGKN-----WFO 563  
DB 597 TSYSTPRLDLIKQAMRYKELFDLKKYNYVITNVTFWGLKDDYSWLSQNFCKSDYPLLED 656  
QY 564 GS-----EPGVTTGGTV----- 580  
DB 657 GNYKSKYAFWSLIEPTVVPVNSTLPAPPAIQVPTPTPTPTPTPTPTPTPTPTPTPTPTPT 716  
QY 581 DGSRFVWAPDGPQVYVAVFGNCSWAASQGVPAQAIRS-----DRVNP 625

DB 717 GGS--YWTPE-----SYGALKVWYANGNMSSTTNVLNPKIKIENVGTTAVDLGRKV 767  
QY 626 KTFVALSNGTFFYRSTGDTGTFQP-----VAAGLPSSCA-----VGMFHAVP 667  
DB 768 RYWTYI-DGEAAQSVSVASSINPAIDVRVVKLGANAGGADYVEVFGKSGAGVLAAGQS 826  
QY 668 QKEGDLMLAASGLYHSTNGSSWSA-----ITGVSSAVNVGFGKAPGSSYPAVFV 719  
DB 827 TKEIRLSIQKSSSYNQNDYSVRGSANSYIENEKVTGYIDDLVLMGREGPGRNAQIKWYA 886  
QY 720 VGTIGGVTA-----YRSDDCGTT-----WVLINDDOHOYGNWQQAITGDHANLR 764  
DB 887 NGNLSSPTNLNPKIKIENVGTTAVDLGRKVRYWYITDGEATQSVSVASSINPAIDVR 946  
QY 765 RYIIGTNGRGIVY-----GDIGCA-----PSGS----- 787  
DB 947 VVKLGANAGGADYVEVFGKSGAGVLAAGQSTKEIRLSIQKSSSYNQNDYSVRGSANSY 1006  
QY 788 -----PSFSVSPASPSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 828  
DB 1007 IENEKVTGYIDDLVLMGKEPSRGTGKAG--EVTAPTPTSTPTPTPTPTPTPTPTPTPTPT 1064  
QY 829 SP 883  
DB 1065 TVTATPT 1124  
QY 884 PCDNIQKPLQVNVNTGSSVDLSVTYVRYWYTRDGGSSLTVYNCWAAICGGINRASFGS 943  
DB 1125 ASTGSIRFWKLVNGSSVDLSRVKIRYWTVDGDKPOSAY-CDWAAQIGASNVTFNFVK 1183  
QY 944 VNPAPTPTADTYLQ 956  
DB 1184 LTSGVSGADYILE 1196

RESULT 15  
ID Q8JLH7 PRELIMINARY; PRT; 196 AA.  
AC Q8JLH7;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Putative endoglucanase (Fragment).  
GN EGL.  
OS Agaricus bisporus (Common mushroom).  
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
OC Agaricales; Agaricaceae; Agaricus.  
OC NCBI\_TaxID=5341;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Morales-Almora P.;  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RA Morales P.;  
RT "Molecular analysis of the cellulolytic genes in Agaricus bisporus.";  
RL Thesis (2001), Department of Microbiology, University of London,  
RL London, United Kingdom.  
DR EMBL; AJ534359; CAD58875.1; --  
FT NON\_TER 1  
FT NON\_TER 196  
SQ SEQUENCE 196 AA; 21353 MW; D9308B2B09935FA7 CRC64;

Query Match 8.0%; Score 409; DB 3; Length 196;  
Best Local Similarity 45.5%; Pred. No. 2.5e-12;  
Matches 76; Conservative 30; Mismatches 51; Indels 10; Gaps 3;

QY 320 KFSVTSGTWTTRISVPSTDTTANDYFGYGLGTLDRQHPNTIMVATQISWMPDTIIFRSTG 379  
DB 1 KYSIDTGAWTDTITPVSGSDL---PFGFGALDITQKGTVMVAALNSWMPDQGIIFRSTNG 57  
QY 380 GATWTRWDWTSYPNRSRYVLDISAEPLWITFGVQPNPPVPSP---KLGWMEAMAIIDPF 436



Search completed: May 14, 2004, 09:18:01  
Job time : 55 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2004, 09:15:08 ; Search time 23 Seconds  
(without alignments)  
2148.089 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MDRSENIRLTWRSRLVSL.....RASFGVNPATPTADTYLOX 957

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/prodata/2/iaa/5A COMB.pep:\*
- 2: /cgn2\_6/prodata/2/iaa/5B COMB.pep:\*
- 3: /cgn2\_6/prodata/2/iaa/6A COMB.pep:\*
- 4: /cgn2\_6/prodata/2/iaa/6B COMB.pep:\*
- 5: /cgn2\_6/prodata/2/iaa/PCTUS COMB.pep:\*
- 6: /cgn2\_6/prodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	334	6.5	1751	3	US-09-136-574A-44	Sequence 44, Appl
2	301.5	5.9	1749	4	US-09-640-419C-28	Sequence 28, Appl
3	288	5.6	96	4	US-09-119-507B-112	Sequence 112, App
4	288	5.6	96	4	US-09-547-693-236	Sequence 236, App
5	276	5.4	1426	3	US-09-136-574A-43	Sequence 43, Appl
6	260	5.1	2736	4	US-09-252-991A-30227	Sequence 30227, A
7	234	4.6	2137	4	US-09-134-001C-4463	Sequence 4463, Ap
8	225.5	4.4	551	2	US-09-033-537A-1	Sequence 1, Appli
9	216.5	4.2	8991	4	US-08-714-741-32	Sequence 32, Appl
10	215.5	4.2	76	4	US-09-547-693-233	Sequence 233, App
11	213	4.1	3892	4	US-09-328-352-5503	Sequence 5503, App
12	210.5	4.1	521	1	US-08-276-213-3	Sequence 3, Appli
13	208.5	4.1	206	4	US-08-529-055-54	Sequence 54, Appl
14	208.5	4.1	493	3	US-09-198-956-10	Sequence 10, Appl
15	208.5	4.1	493	3	US-09-198-955A-12	Sequence 12, Appl
16	208.5	4.1	493	4	US-09-694-531-12	Sequence 12, Appl
17	208.5	4.1	493	4	US-09-670-141-10	Sequence 10, Appl
18	208.5	4.1	493	4	US-10-072-152-12	Sequence 12, Appl
19	205	4.0	918	4	US-09-200-650B-1	Sequence 1, Appli
20	202.5	3.9	476	4	US-09-339-159B-4	Sequence 4, Appli
21	200.5	3.9	700	2	US-07-862-588B-2	Sequence 2, Appli
22	200	3.9	183	4	US-08-529-055-50	Sequence 50, Appl
23	199.5	3.9	490	3	US-09-109-841-2	Sequence 2, Appli
24	199.5	3.9	616	3	US-09-136-574A-47	Sequence 47, Appl
25	197	3.8	1060	4	US-08-911-393-2	Sequence 2, Appli
26	191.5	3.7	412	1	US-08-313-288B-18	Sequence 18, Appl
27	191	3.7	105	4	US-09-547-693-230	Sequence 230, App

28	189	3.7	933	3	US-08-293-728-2	Sequence 2, Appli
29	189	3.7	933	3	US-09-421-868-2	Sequence 2, Appli
30	188.5	3.7	193	4	US-08-529-055-49	Sequence 49, Appl
31	187	3.6	936	4	US-08-956-171E-5249	Sequence 5249, Ap
32	186.5	3.6	104	4	US-09-547-693-235	Sequence 235, App
33	186	3.6	423	2	US-08-760-797A-1	Sequence 1, Appli
34	185.5	3.6	424	3	US-08-932-929B-1	Sequence 1, Appli
35	185.5	3.6	1719	2	US-08-459-568-4	Sequence 4, Appli
36	185.5	3.6	1719	2	US-08-399-411-4	Sequence 4, Appli
37	185.5	3.6	1719	3	US-08-516-859A-4	Sequence 4, Appli
38	185.5	3.6	1719	4	US-09-586-472-4	Sequence 4, Appli
39	185.5	3.6	1719	4	US-09-528-706-4	Sequence 4, Appli
40	185	3.6	334	6	5202236-3	Patent No. 5202236
41	184	3.6	331	6	5202236-37	Patent No. 5202236
42	181	3.5	167	5	PCT-US95-13813-9	Sequence 9, Appli
43	180.5	3.5	424	2	US-08-760-797A-3	Sequence 3, Appli
44	180.5	3.5	424	3	US-08-932-929B-3	Sequence 3, Appli
45	179.5	3.5	1481	2	US-08-616-844-40	Sequence 40, Appl

ALIGNMENTS

RESULT 1

US-09-136-574A-44  
; Sequence 44, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for

; Treating Cellulose Containing Fabrics Using Truncated

; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/932,571

; FILING DATE: September 19, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Bak, Mary E.

; REGISTRATION NUMBER: 31,215

; REFERENCE/DOCKET NUMBER: 1997US001/CIP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-540-9200

; TELEFAX: 215-540-5818

; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 44:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1751 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 44:



```
Db 503 -GNLTFSPTITIDALFS-----LGTGSIIPANTGCPINTGNTLVTSPTNLDNDKIQIQLNA 557
Qy 467 -----APMWKGBLETAVN-----DLISPPS-----GAPLIISALGDLG----- 498
Db 558 NTFPTAFGYDIYPRKSVNGGFULLGNANFVYFDDGAGITTPKSGFVGTATEIYVGNDDIVT 617
Qy 499 -----GTHADVT-AVPSITFTSPVFTTG----- 617
Db 618 QGNQIFDGVFYGLQPNVLTSSAGSVITNNILLGSLQVTAQNIIVSQSSLSAVEIAS 677
Qy 532 SILVRAGSPDQPNDRHAF---STDGKNWFGQSGPGVTTG-----GTVA 577
Db 678 DVLNAG-----QNVSGFNTRGNNVDIQLA--GNISTGSIVTSFPGNAGNVI 725
Qy 578 ASADG---SRFWAPGDPGPVVYAVFGNSWA-----ASQGV 612
Db 726 LNAAGTITTCYIETSGTNGGDDVTSSGNTSTAVIDTRGFGDGLDLSLGGAVSIESKGD 785
Qy 613 PANAIQISDRVNPKTFYVALNGTFTYRSTGCVTPQPVAAAGLPSSGAVGVMFHVPKGE-- 670
Db 786 ITTAFIDTGAYSTESFNEGTVGNVFLTAGDSITTYITAGKNGG--DIFFOAGESIEII 843
Qy 671 -----GDLWLA--SSGLYHSTNGGS-----SHSAITGV 697
Db 844 DYLNTYGSQTSQDVYVEAPLDISIGSYITGCGGEGPQNVFLQAGGDITTSYIDTSAANGG 903
Qy 698 SSANVVG-----FGKSAPGSSYPVAVFVVG 722
Db 904 DIFQSGDTEVGVLFYKGVGEGRGDVYVETGRYFRAIDGFLGEEGFFSVYTAGLVG- 962
Qy 723 IGVGTGAYRSDDCGTTWVLINDHQYGNMGQAITGDHANLRRIYIGTNGR--GIVYVGD 780
Db 963 -----GSVVI-----QFGSGEPFIIGNPI-----TNGTIGALSSGDD 994
Qy 781 GGAPSGSP-----SPSVSPASPSLSPPSSPSSPSSPSSPSSPSSPSSPSSPSSPSS 825
Db 995 NTVPIGTPIFDFTLNDITITTEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE 1054
Qy 826 PPSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 868
Db 1055 PEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE 1097

RESULT 3
US-09-119-507B-112
; Sequence 112, Application US/09119507B
; Patent No. 6548642
; GENERAL INFORMATION:
; APPLICANT: Kieliszewski, Marcia J.
; TITLE OF INVENTION: No. 6548642el Synthetic Genes for Plant Gums
; FILE REFERENCE: OHU-03417
; CURRENT APPLICATION NUMBER: US/09/119,507B
; CURRENT FILING DATE: 1998-07-20
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 112
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-119-507B-112
Query Match 5.6%; Score 288; DB 4; Length 96;
Best Local Similarity 77.0%; Pred. No. 1.1e-11;
Matches 57; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

Qy 791 SVSPASPSLSPPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSP 850
Db 20 SLSLAQTTRASPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 79
Qy 851 SPSSSPSSSPSPPTP 864
Db 80 SPSPSPSPSPSPSP 93

RESULT 5
US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
```

```
Db 80 SPSPSPSPSPSPSP 93

RESULT 4
US-09-547-693-236
; Sequence 236, Application US/09547693
; Patent No. 6639050
; GENERAL INFORMATION:
; APPLICANT: Kieliszewski, Marcia
; TITLE OF INVENTION: Synthetic Genes for Plant Gums and Other Hydroxyproline-Rich
; Glycoproteins
; FILE REFERENCE: OHU-04089
; CURRENT APPLICATION NUMBER: US/09/547,693
; CURRENT FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 236
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 236
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Artificial/Unknown
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Synthetic
US-09-547-693-236
Query Match 5.6%; Score 288; DB 4; Length 96;
Best Local Similarity 77.0%; Pred. No. 1.1e-11;
Matches 57; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

Qy 791 SVSPASPSLSPPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSPSSP 850
Db 20 SLSLAQTTRASPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 79
Qy 851 SPSSSPSSSPSPPTP 864
Db 80 SPSPSPSPSPSPSP 93

RESULT 5
US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
```



US-09-134-001C-4463  
; Sequence 4463, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4463  
; LENGTH: 2137  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4463

Query Match 4.6%; Score 234; DB 4; Length 2137;  
Best Local Similarity 22.0%; Pred. No. 1.9e-06;  
Matches 214; Conservative 129; Mismatches 404; Indels 226; Gaps 41;

QY 59 IGGGFGVDFGIVNEGAPGLYVRTDIGGMYRWDAAAGRWI-----PLLDWYGNWNGY 111  
Db 435 VGGGAF-----GA----FVSTDRNGMATTEASSAAKLNVQPTDNSQDFV----IDY 479  
QY 112 NG-----VVSIAADPINTNKWAAVGMYNWDPNDGAILRSSDQATWQITPLPKLGG 166  
Db 480 NGDTKMTVTYAGQTFRN-----LTDWIKNSG-----GTFPSLS-MTASTGG 521  
QY 167 --NMPGRGGERLAVDPNDNINLYFCAPSKGL-----197  
Db 522 AKNLOQVQGTFBYTESAVAKRYVDANTGKOIIPKTIAGEVDATVNDKQLNNLKN 581  
QY 198 --WRSTDS--GATWSQMTNPPDVGTVIANTPTDTGYQ--SDIQGVWVAFDKSSSLGQAS 252  
Db 582 YSVSTDALQNSYSTSGTPTL--KLTHSSQTVIYKFDVQV--PQISVDSDQTRVGVKTI 638  
QY 253 KTFVGVADPNPNVFWRSRGGATWQAVPGAPTGFIPHKGVDPVNVHLYIATNTGPGYD 312  
Db 639 NPITITTDNSKDLVLT-----IVTGLPSGL-----SFDQTWTI-----TGTPSE 679  
QY 313 -----GSSGVWKFVSCTGWTTRISPV-----PSTDTANDYFG 345  
Db 680 VGTTVTVNTDAGTGNVTSKQFTITTIQDTISPVVNVTPSQASEVFTPIINPITITATD 739  
QY 346 Y-----SGLTIDRQHPNTIMVATOISWPDITIFRSTP--GGATWTRI--WDWT-- 390  
Db 740 KVTHVTGLPQGLKFDASTNSIVGPTQIG--TNTITITSDAGNKTKIKINYEVRN 797  
QY 391 --SYPNRSLRYV-----LDISAEPLWTFGVQPNPVPSPKLGWMDMAIDPFNSDRMLYGT 445  
Db 798 SASDSTSTSVNSVSTSIKSTSLSDSVKASQSLSTKS--LSELSASTNSNSTSIQASE 855  
QY 446 GAT-----LVATNDLTKWDSGGQTHIAPWVKLBETAVNDLISPPSGAPLISALGDLGGFT 501  
Db 856 SASSTKQLESASTSDSASE-----SARKSESTSKSTLSSESTSTSVSDSASVSTSES 910  
QY 502 HADTVAVPTIFTSVPTTGTSTVDYAEALNPSIIVRAGSPDPSPQPNDRHVAFTDGGKNW 561  
Db 911 ASTSTSVSGSTSTSIKSTSTST-----SDSASIKASESASTSKLSESVSTST----- 959  
QY 562 FQSGEPGVTTGGTVAASADGSRFVWAPDGPQPVVYVAVFGNSWAASQCVPANAIQRSD 621  
Db 960 ---SDSASTSTSVSDNSASTSL-----SKSTSTSVSDSTSTSTSDSASTST---SE 1005  
QY 622 RVNPKTFYALNGTFFRSTDTGGVTFQVPAAGLPSGAGVGMFHVAPGKGDMLWAASGL 681  
Db 1006 SEDSDASTSLSESTSTSVSDSTSTSTSDSASMSASESES-----NSKSTSLSESTSTSL 1059

QY 682 YHSTNGSSMSAITGVSSAVNVGFGKAPGSSYPAVFVVGTIGGVTGAYRSDDCGTWVL 741  
Db 1060 SGSTASSTSDSASTSTSESDSTSTSLSESTST-----SLSGSTASSTSDSASTS-TS 1112  
QY 742 INDDQHOYGNWQAI-----TGDHANLRVYIGTNGRGI--VYGDIGCAPSGSPS 789  
Db 1113 ESUSTSESTSLSESLSTSVSDSTASTSESTSESTSESTSESTSESTSESTSLSDSTS 1172  
QY 790 PSVPSASPLSPSS 849  
Db 1173 TSTSDASSTSTSESDSTSTSLSESTSTSLSDSTSTSTSESTSESTSESTSESTSL 1232  
QY 850 PSPSSSP 909  
Db 1233 ESTSTSVSDSTASTSDS--ASTSTSV-----SDSEASTSISESL-----STSVSDSTST 1281  
QY 910 VRYWFTREDGGSST 922  
Db 1282 ----STSDSASTS 1290

RESULT 8  
US-09-033-537A-1  
; Sequence 1, Application US/09033537A  
; Patent No. 5958083  
; GENERAL INFORMATION:  
; APPLICANT: Onishi, Masahiro  
; APPLICANT: Fich, Merete  
; APPLICANT: Tof, Annette Hanne  
; APPLICANT: Sh lein, Martin  
; TITLE OF INVENTION: Prevention Of Back-Staining  
; TITLE OF INVENTION: In Stone Washing  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: U.S.A.  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA: US/09/033,537A  
; APPLICATION NUMBER: 0993/95  
; FILING DATE: 08-SEP-1995  
; APPLICATION NUMBER: PCT/DK96/00364  
; FILING DATE: 03-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Green, Reza  
; REGISTRATION NUMBER: 38,475  
; REFERENCE/DOCKET NUMBER: 4492.204-US  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 551 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-033-537A-1

Query Match 4.4%; Score 225.5; DB 2; Length 551;  
Best Local Similarity 23.8%; Pred. No. 1.1e-06;  
Matches 114; Conservative 72; Mismatches 166; Indels 127; Gaps 22;







41	DB	LWSRDYRSLMDQIKS--LGYNTRLRPSYDDILKPGTWPNSINFYQMNQDLQGLTSLQVMD	98
481	QY	LISPSPGAPLISALGDIG---GFTHADVTPSTIFTSPVFTTGTSDVAEL-----	529
99	DB	KI-----VAVAGQIGLRILIIIDRRHPCSCGQALWYTSVSSEATWISDLQALAQRYKG	150
530	QY	NPSIIVRAGSFDPSSOPNDRHVAFTSDGKXNWFQSGEPGG-----VTTGGTVAA	578
151	DB	NETVV---GFDLNHEPHDPACWCGGDPISDWRLAAERAGNAVLNVNPNLLIFVEG--VQ	204
579	QY	SADGSRFVWAPGDPGPQPVVYAVGFGNSWAASQ-----GVP---ANAQIRSDRVNPKTFY	629
205	DB	SYNGDSYWG-----GNLQAGQYPVVLNVPNRLVYSAHDYATSVVPQTFW	250
630	QY	ALSNGTFYRSTDCGVTFPPVAAGLPSSGAGVGMFHAVPGKEGDIWLAASSGLYHSTNGGS	689
251	DB	--SDPTF-----PNN-----MPGIMNKNW-----GYLFNQNIAP	277
690	QY	SWSAITG--VSSAVNVGFGKSAPGSSYPAPVVGVTIGGVTGAYRSD-----DCG	736
278	DB	VMLGEFGTLLQSTTDQTLWLKTLVQYLRP-----TAQYGADSPQWTFWGNPDSG	326
737	QY	TTWVLINDHOHGXNMGQAITGDHANLRVYVIGTNGRGIYVDGIGAPSGSPSPSVSPA	796
327	DB	DTGGILKDD-----WQVDTVKDGYLAPI-----KSSIFDPVGA-----SA	362
797	QY	SPSLSP	856
363	DB	SP--SQPSP	418
857	QY	SSSPSPTPSPSPSPSGGVKVQYK--NNDSPAGDNQIKPGQLQVNTGSSVDLSTVTYVWFT	915
419	DB	ASGARCTAS-----YQVNSDMGNGFT-----VTVAVTNMSGVATKTTWTVSWTF-	461
916	QY	RDGGSSTLVNCDWAA	931
462	DB	--CGNOTITNS--WNA	473

RESULT 13
US-08-529-055-54
; Sequence 54, Application US/08529055
; Patent No. 6592876
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yoether, Janet
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: Pneumococcal Genes, Portions
; THEREOF: Expression Products
; TITLE OF INVENTION: Thereof, Expression Products
; TITLE OF INVENTION: Therefrom, and Uses of Such Genes,
; TITLE OF INVENTION: Portions and Products
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/529,055
; FILING DATE: 15-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.

;	CURRENT APPLICATION NUMBER: US/09/198,956
;	CURRENT FILING DATE: 1998-11-24
;	EARLIER APPLICATION NUMBER: 1344/97
;	EARLIER FILING DATE: 1997-11-24
;	EARLIER APPLICATION NUMBER: 60/067,240
;	EARLIER FILING DATE: 1997-12-02
;	NUMBER OF SEQ ID NOS: 26
;	SOFTWARE: FastSeq for Windows Version 3.0
;	SEQ ID NO 10
;	LENGTH: 493
;	TYPE: PRF
;	ORGANISM: Bacillus licheniformis
;	US-09-198-956-10

Query Match	4.1%; Score 208.5; DB 3; Length 493;
Best Local Similarity	23.5%; Pred. No. 1.2e-05; Indels 125; Gaps 19;
Matches	103; Conservative 62; Mismatches 149;

QY	568	GVTTCGTVAASAD----	GSRFVWAPGDPQPVVYAVGVFG--NSWAASQGVPANAQIRSDR	622
Db	63	GTITTSNTSASKIDVKDVSNVSIVSGTKGE---	LKGIGIKIWRANNIIIRNLKIH-----	116
QY	623	VNPKTFFALNGTFYRSTGGVTTFQPVAAAGLPSSGAUGVMFHAVPGEGDGLWLASSGLY	682	
Db	117	-----VASG--DKDAIG-----IEGPSKNIWVDHNE-LY	142	
QY	683	HSTNCGSSW-----SAITGVSSAVNVVGFKSPAGSVPAVFVGTIGGVTGAYR	731	
Db	143	HSLNVDKYDDGLFDVKRDAEYLTFWSNYVDHGKSKMLMGSS-----	185	
QY	732	SDCGTTWWVLIND-----DOHQYNWGQAITGDHANLR---RVVIGNGR	773	
Db	186	SDNNRTTIFPHHNPFNLNRVPSFPFEIGHIYYNYFNKIIDSINSRMGARIRINN--	243	

QY 774 GIVYDGGAPSGSPSPVS-PSASPS---LSPSP---SPSSSPSPSPSPSPSPSPSP 826  
Db 244 -----LFEAKDPIVSWYSSPGYKWSNKNKFNNSRGSMPTTSTTTTYPYSYSLD 294  
QY 827 SP 886  
Db 295 NVDNKSIVKQ---NAGVKIQRRPPTPTSPSPSPSPSPSPSPSPSPSPSPSPSP 348  
QY 887 NQIKPGLOVNTGSSVDLSVTVRYWFTRDGSSSTLVYNCDAWAAI-----GCGNI 937  
Db 349 NSINPQKVTNTGSSAIDLSKLTLYYYTVGQKQDTFW-CDHAAIIGSNGSYNGITSNV 407  
QY 938 RASFGSVNPATPTADTYLQ 956  
Db 408 KGTfVMSSTNNADTYLE 426

## RESULT 15

US-09-198-955A-12  
; Sequence 12, Application US/09198955A  
; Patent No. 6187580  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Lene N.  
; APPLICANT: Schuelein, Martin  
; APPLICANT: Lange, Niels E.  
; APPLICANT: Bjornvad, Mads E.  
; APPLICANT: Moller, Soren  
; APPLICANT: Glad, Sanne O. S.  
; APPLICANT: Kauppinen, Markus S.  
; APPLICANT: Schnorr, Kirk  
; APPLICANT: Kongsbak, Lars  
; TITLE OF INVENTION: No. 6187580el Pectate Lyases  
; FILE REFERENCE: 5378.200-US  
; CURRENT APPLICATION NUMBER: US/09/198, 955A  
; CURRENT FILING DATE: 1998-11-24  
; PRIOR APPLICATION NUMBER: 1343/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 1344/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/067,249  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 60/067,240  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 09/073,684  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: 09/184,217  
; PRIOR FILING DATE: 1998-11-02  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 493  
; TYPE: PRT  
; ORGANISM: Clostridium thermocellum  
US-09-198-955A-12

Query Match 4.1%; Score 208.5; DB 3; Length 493;  
Best Local Similarity 23.5%; Pred. No. 1.2e-05;  
Matches 103; Conservative 62; Mismatches 149; Indels 125; Gaps 19;  
QY 568 GGVTTGGTVAASAD----GSRFVWAPGDPQPPVYAVGFG-NSWAAASQGVPAQAIRSDR 622  
Db 63 GTTTSNTSASKIDVKDVSNSIVSGTGKE-----LKGIGIKIWRANNIIRNLKHE-- 116  
QY 623 VNPKTFYALSNGTFRSTGGVTFQPVAAGLPSSGAGVGMFHAVPGKEGDLWLAASSGLY 682  
Db 117 -----VASG--DKDAIG-----IEGPKNIWVDHNE-LY 142  
QY 683 HSTNGSSW-----SAITGVSSAVNVGFGKAPGSSYPVAVVGTIGVGTGAYR 731  
Db 143 HSLNVDKDYDGLDFVKRDAEYITFSWYVHDGWSKMLMGSS-----D 185  
QY 732 SDDCGTTWLIND-----DQHQYGNWQAITGDHANLR---RVYIGTNGR 773

Db 186 SDYNRTITTFHHNWFENLSRVSPFRFGEGHYNNYFNKLIIDSGINSRMGARIRIENN-- 243  
QY 774 GIVYDGGAPSGSPSPVS-PSASPS---LSPSP---SPSSSPSPSPSPSPSPSPSP 826  
Db 244 -----LFEAKDPIVSWYSSPGYKWSNKNKFNNSRGSMPTTSTTTTYPYSYSLD 294  
QY 827 SP 886  
Db 295 NVDNKSIVKQ---NAGVKIQRRPPTPTSPSPSPSPSPSPSPSPSPSPSPSPSP 348  
QY 887 NQIKPGLOVNTGSSVDLSVTVRYWFTRDGSSSTLVYNCDAWAAI-----GCGNI 937  
Db 349 NSINPQKVTNTGSSAIDLSKLTLYYYTVGQKQDTFW-CDHAAIIGSNGSYNGITSNV 407  
QY 938 RASFGSVNPATPTADTYLQ 956  
Db 408 KGTfVMSSTNNADTYLE 426

Search completed: May 14, 2004, 09:19:15  
Job time : 27 secs

Qy	1	MORSENI	RLT	WRSRRL	SVLSLL	LAATAS	FAVAAL	GLVPL	PIAITAS	PAHAA	TTQPT	WTSN	VNVAIG	60	
Db	1	MORSENI	RLT	WRSRRL	SVLSLL	LAATAS	FAVAAL	GLVPL	PIAITAS	PAHAA	TTQPT	WTSN	VNVAIG	60	
Qy	61	GGG	VDG	IVF	NFEG	AGIL	YVR	TDI	IGGMY	RWDA	ANGRWI	PLLD	WGWNN	NGVGVYSIAAD	12
Db	61	GGG	VDG	IVF	NFEG	AGIL	YVR	TDI	IGGMY	RWDA	ANGRWI	PLLD	WGWNN	NGVGVYSIAAD	12

QY 121 PINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPFKLGNNPGRGMRGLAVD 180  
Db 121 PINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPFKLGNNPGRGMRGLAVD 180  
QY 181 PNNDNLLYFGAPSGKGLWSTDSGATWSQMTNPDVCTYIANPTDTTGYQSDIQGVVWVA 240  
Db 181 PNNDNLLYFGAPSGKGLWSTDSGATWSQMTNPDVCTYIANPTDTTGYQSDIQGVVWVA 240  
QY 241 FDKSSSLGQASKTIFVGVADPNPNVFWSDRGATWQAVPGAPTGFPHKGVDPVNVHL 300  
Db 241 FDKSSSLGQASKTIFVGVADPNPNVFWSDRGATWQAVPGAPTGFPHKGVDPVNVHL 300  
QY 301 YIATNTGGPYDSSGSDVWKFSTGTRISPVSTDTTANDYFGYSGLTIDRQHPNTIM 360  
Db 301 YIATNTGGPYDSSGSDVWKFSTGTRISPVSTDTTANDYFGYSGLTIDRQHPNTIM 360  
QY 361 VATQISWMPDTIIFRSTDDGATWTRIDWNTSYPNRSRYVLDISAEPLWTFGVQNPVP 420  
Db 361 VATQISWMPDTIIFRSTDDGATWTRIDWNTSYPNRSRYVLDISAEPLWTFGVQNPVP 420  
QY 421 SPKLGWMDDEAMADPPNSDRMLYGTGATLYATNDLTWDSGGQIHIAPMVKGLEETA VND 480  
Db 421 SPKLGWMDDEAMADPPNSDRMLYGTGATLYATNDLTWDSGGQIHIAPMVKGLEETA VND 480  
QY 481 LISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSVFTTGTSDVYAEINPSIIVRAGSF 540  
Db 481 LISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSVFTTGTSDVYAEINPSIIVRAGSF 540  
QY 541 DPSSQPNDRHVAESTDGGKWNFGQSEPGVTTGGTVAASADGSRFVWAPGDPQPVVYAV 600  
Db 541 DPSSQPNDRHVAESTDGGKWNFGQSEPGVTTGGTVAASADGSRFVWAPGDPQPVVYAV 600  
QY 601 GFNSWAASQGVVPANAQIRSDRVNPKTFYALSNGTFRSTGGVTFQPVAAAGLPSSGAVG 660  
Db 601 GFNSWAASQGVVPANAQIRSDRVNPKTFYALSNGTFRSTGGVTFQPVAAAGLPSSGAVG 660  
QY 661 VMFHAVPGEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFGKSAFGSSYPVAVV 720  
Db 661 VMFHAVPGEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFGKSAFGSSYPVAVV 720  
QY 721 GTTGGVTA YRSDDCGTTWVLINDDHOYGNWQAITGDHANLRVYIGTNGRGIVYGD I 780  
Db 721 GTTGGVTA YRSDDCGTTWVLINDDHOYGNWQAITGDHANLRVYIGTNGRGIVYGD I 780  
QY 781 GGAPSGSPSVSP 840  
Db 781 GGAPSGSPSVSP 840  
QY 841 SASP 900  
Db 841 SASP 900

RESULT 2

US-10-155-400-1  
; Sequence 1, Application US/10155400  
; Publication No. US2003010898A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE OF INVENTION: CELLULOYTICUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid  
US-10-155-400-1  
  
Query Match 100.0%; Score 5134; DB 14; Length 957;  
Best Local Similarity 100.0%; Pred. No. 1e-290;  
Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MDRSENRLTMRSRRLVSLAATAFVAALALGVLPITAITASPAHAATQPTYSNVAIG 60  
Db 1 MDRSENRLTMRSRRLVSLAATAFVAALALGVLPITAITASPAHAATQPTYSNVAIG 60  
QY 61 GGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWITPLLDWGNWNGYNGVVSIAAD 120  
Db 61 GGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWITPLLDWGNWNGYNGVVSIAAD 120  
QY 121 PINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPFKLGNNPGRGMRGLAVD 180  
Db 121 PINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPFKLGNNPGRGMRGLAVD 180  
QY 181 PNNDNLLYFGAPSGKGLWSTDSGATWSQMTNPDVCTYIANPTDTTGYQSDIQGVVWVA 240  
Db 181 PNNDNLLYFGAPSGKGLWSTDSGATWSQMTNPDVCTYIANPTDTTGYQSDIQGVVWVA 240  
QY 241 FDKSSSLGQASKTIFVGVADPNPNVFWSDRGATWQAVPGAPTGFPHKGVDPVNVHL 300  
Db 241 FDKSSSLGQASKTIFVGVADPNPNVFWSDRGATWQAVPGAPTGFPHKGVDPVNVHL 300  
QY 301 YIATNTGGPYDSSGSDVWKFSTGTRISPVSTDTTANDYFGYSGLTIDRQHPNTIM 360  
Db 301 YIATNTGGPYDSSGSDVWKFSTGTRISPVSTDTTANDYFGYSGLTIDRQHPNTIM 360  
QY 361 VATQISWMPDTIIFRSTDDGATWTRIDWNTSYPNRSRYVLDISAEPLWTFGVQNPVP 420  
Db 361 VATQISWMPDTIIFRSTDDGATWTRIDWNTSYPNRSRYVLDISAEPLWTFGVQNPVP 420  
QY 421 SPKLGWMDDEAMADPPNSDRMLYGTGATLYATNDLTWDSGGQIHIAPMVKGLEETA VND 480  
Db 421 SPKLGWMDDEAMADPPNSDRMLYGTGATLYATNDLTWDSGGQIHIAPMVKGLEETA VND 480  
QY 481 LISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSVFTTGTSDVYAEINPSIIVRAGSF 540  
Db 481 LISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSVFTTGTSDVYAEINPSIIVRAGSF 540  
QY 541 DPSSQPNDRHVAESTDGGKWNFGQSEPGVTTGGTVAASADGSRFVWAPGDPQPVVYAV 600  
Db 541 DPSSQPNDRHVAESTDGGKWNFGQSEPGVTTGGTVAASADGSRFVWAPGDPQPVVYAV 600  
QY 601 GFNSWAASQGVVPANAQIRSDRVNPKTFYALSNGTFRSTGGVTFQPVAAAGLPSSGAVG 660  
Db 601 GFNSWAASQGVVPANAQIRSDRVNPKTFYALSNGTFRSTGGVTFQPVAAAGLPSSGAVG 660  
QY 661 VMFHAVPGEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFGKSAFGSSYPVAVV 720  
Db 661 VMFHAVPGEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFGKSAFGSSYPVAVV 720  
QY 721 GTTGGVTA YRSDDCGTTWVLINDDHOYGNWQAITGDHANLRVYIGTNGRGIVYGD I 780  
Db 721 GTTGGVTA YRSDDCGTTWVLINDDHOYGNWQAITGDHANLRVYIGTNGRGIVYGD I 780  
QY 781 GGAPSGSPSVSP 840  
Db 781 GGAPSGSPSVSP 840  
QY 841 SASP 900  
Db 841 SASP 900

Qy 901 SSVDLSTVTVRYWFTRDGSSSTLVNCDMAAIGCGNIRASFGSVNPAITPDTATYQ 956  
|  
Db 901 SSVDLSTVTVRYWFTRDGSSSTLVNCDMAAIGCGNIRASFGSVNPAITPDTATYQ 956

## RESULT 3

US-09-917-376-3  
; Sequence 3, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; OTHER INFORMATION: Catalytic domain GH74

US-09-917-376-3

Query Match 78.6%; Score 4036; DB 12; Length 740;  
Best Local Similarity 100.0%; Pred. No. 6.9e-227;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106  
|  
Db 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
|  
Qy 107 NNWYNGVWSIAADPINTNKVAAVGMVYNSWPNDCAILRSSDQATWQITPLPKLGG 166  
|  
Db 61 NNWYNGVWSIAADPINTNKVAAVGMVYNSWPNDCAILRSSDQATWQITPLPKLGG 120  
|  
Qy 167 NMPGRGGERLAVDPNNNDILYFGAPSGKGLWSTDSGATWSQMTNFPDVGTYIANPTDT 226  
|  
Db 121 NMPGRGGERLAVDPNNNDILYFGAPSGKGLWSTDSGATWSQMTNFPDVGTYIANPTDT 180  
|  
Qy 227 TGYQSDIQQVYVWVAFDKSSSLGQASKTIIFGVADPNPNPVFWSRDGGATWQAVPGAPTGF 286  
|  
Db 181 TGYQSDIQQVYVWVAFDKSSSLGQASKTIIFGVADPNPNPVFWSRDGGATWQAVPGAPTGF 240  
|  
Qy 287 IPHKGVPDPVNHVLYIATSNITGGPYDGGSDGVKFSVTSCTWTRISVPVSTDTANDYFGY 346  
|  
Db 241 IPHKGVPDPVNHVLYIATSNITGGPYDGGSDGVKFSVTSCTWTRISVPVSTDTANDYFGY 300  
|  
Qy 347 SGLTIDRQHPTIMVATQISWMPDITIIFRSTDDGATWTRIDWTSYPNRSRLRYVLDISAE 406  
|  
Db 301 SGLTIDRQHPTIMVATQISWMPDITIIFRSTDDGATWTRIDWTSYPNRSRLRYVLDISAE 360  
|  
Qy 407 PMLTFGQVQPNPPVPSPKLGMWDEAMAIIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 466  
|  
Db 361 PMLTFGQVQPNPPVPSPKLGMWDEAMAIIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
|  
Qy 467 APVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIPTSVPFTTGTSDY 526  
|  
Db 421 APVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIPTSVPFTTGTSDY 480  
|  
Qy 527 AELNPSIIVRAGSFDPSQPNDRHVAFTDGGKWNFGSEPGVGTGCTVAASADGSRFV 586  
|  
Db 481 AELNPSIIVRAGSFDPSQPNDRHVAFTDGGKWNFGSEPGVGTGCTVAASADGSRFV 540  
|  
Qy 587 WAPGDPQCPVYVAVGFGNSWAASQGVPANAQIRSDRWNPKTFYALSNGTIFYRSTDDGVTF 646  
|  
Db 541 WAPGDPQCPVYVAVGFGNSWAASQGVPANAQIRSDRWNPKTFYALSNGTIFYRSTDDGVTF 600  
|

Qy 647 QPVAAGLPSSGAVGVMFHVPKKEGDLMLAASSGLYHSTNGSSWSAITGVSAVNVGFG 706  
|  
Db 601 QPVAAGLPSSGAVGVMFHVPKKEGDLMLAASSGLYHSTNGSSWSAITGVSAVNVGFG 660  
|  
Qy 707 KSAFGSSYPVAVVGTIGCVTGAYRSDCCGTWVLINDDOHQYGNWQAITGDHANLRRV 766  
|  
Db 661 KSAFGSSYPVAVVGTIGCVTGAYRSDCCGTWVLINDDOHQYGNWQAITGDHANLRRV 720  
|  
Qy 767 YIGTNGRGIVYGDIGGAPSG 786  
|  
Db 721 YIGTNGRGIVYGDIGGAPSG 740  
|

## RESULT 4

US-09-917-376-6  
; Sequence 6, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus

US-09-917-376-6

Query Match 78.6%; Score 4036; DB 12; Length 740;  
Best Local Similarity 100.0%; Pred. No. 6.9e-227;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106  
|  
Db 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
|  
Qy 107 NNWYNGVWSIAADPINTNKVAAVGMVYNSWPNDCAILRSSDQATWQITPLPKLGG 166  
|  
Db 61 NNWYNGVWSIAADPINTNKVAAVGMVYNSWPNDCAILRSSDQATWQITPLPKLGG 120  
|  
Qy 167 NMPGRGGERLAVDPNNNDILYFGAPSGKGLWSTDSGATWSQMTNFPDVGTYIANPTDT 226  
|  
Db 121 NMPGRGGERLAVDPNNNDILYFGAPSGKGLWSTDSGATWSQMTNFPDVGTYIANPTDT 180  
|  
Qy 227 TGYQSDIQQVYVWVAFDKSSSLGQASKTIIFGVADPNPNPVFWSRDGGATWQAVPGAPTGF 286  
|  
Db 181 TGYQSDIQQVYVWVAFDKSSSLGQASKTIIFGVADPNPNPVFWSRDGGATWQAVPGAPTGF 240  
|  
Qy 287 IPHKGVPDPVNHVLYIATSNITGGPYDGGSDGVKFSVTSCTWTRISVPVSTDTANDYFGY 346  
|  
Db 241 IPHKGVPDPVNHVLYIATSNITGGPYDGGSDGVKFSVTSCTWTRISVPVSTDTANDYFGY 300  
|  
Qy 347 SGLTIDRQHPTIMVATQISWMPDITIIFRSTDDGATWTRIDWTSYPNRSRLRYVLDISAE 406  
|  
Db 301 SGLTIDRQHPTIMVATQISWMPDITIIFRSTDDGATWTRIDWTSYPNRSRLRYVLDISAE 360  
|  
Qy 407 PMLTFGQVQPNPPVPSPKLGMWDEAMAIIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 466  
|  
Db 361 PMLTFGQVQPNPPVPSPKLGMWDEAMAIIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
|  
Qy 467 APVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIPTSVPFTTGTSDY 526  
|  
Db 421 APVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIPTSVPFTTGTSDY 480  
|  
Qy 527 AELNPSIIVRAGSFDPSQPNDRHVAFTDGGKWNFGSEPGVGTGCTVAASADGSRFV 586  
|

Db 481 AELNPSIIVRAGSFQSSQPNDRHVAFASTDDGKQWFGSEPGVTTGGTVAASADGSRFV 540  
QY 587 WAPGDPQPVVAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALNGTFYRSTDDGVT 646  
Db 541 WAPGDPQPVVAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALNGTFYRSTDDGVT 600  
QY 647 QPVAAGLPSSGAVGVMFHVPKGGDLWLAASSGLYHSTNGSSWSAITGVSAVNVGFG 706  
Db 601 QPVAAGLPSSGAVGVMFHVPKGGDLWLAASSGLYHSTNGSSWSAITGVSAVNVGFG 660  
QY 707 KSAFGSSYPVAVVGTIGVGTGAYRSDDCGTTWVLINDDQHOYGNWQOAITGDHANLRV 766  
Db 661 KSAFGSSYPVAVVGTIGVGTGAYRSDDCGTTWVLINDDQHOYGNWQOAITGDHANLRV 720  
QY 767 YIGTNGRGIVYGDIGGAPSG 786  
Db 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 5  
US-10-155-400-3  
; Sequence 3, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; OTHER INFORMATION: Catalytic domain GH74  
US-10-155-400-3

Query Match 78.6%; Score 4036; DB 14; Length 740;  
Best Local Similarity 100.0%; Pred. No. 6.9e-227;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 ATTPQYTWNSNVAIGGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWVGW 106  
Db 1 ATTPQYTWNSNVAIGGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWVGW 60  
QY 107 NNWYNGVWSIAADPINTNKWAAVGMVYNSWDPNDGAILRSSDQGTATWQITPLPFKLG 166  
Db 61 NNWYNGVWSIAADPINTNKWAAVGMVYNSWDPNDGAILRSSDQGTATWQITPLPFKLG 120  
QY 167 NMPGRGMLERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNPPDVGTYIANPTDT 226  
Db 121 NMPGRGMLERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNPPDVGTYIANPTDT 180  
QY 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNNPVFWSRSDGATWQAVPGATGF 286  
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNNPVFWSRSDGATWQAVPGATGF 240  
QY 287 IPHKGVDPVNVHLYIATSNTPGPDGSGDVWKFVSTGTWTRISPVSTDTANDYFGY 346  
Db 241 IPHKGVDPVNVHLYIATSNTPGPDGSGDVWKFVSTGTWTRISPVSTDTANDYFGY 300  
QY 347 SGLTIDROHNTIMVATQISWNPDTIIFRSTDDGATWTRIDWTSYPNRSLRYVLDISAE 406  
Db 301 SGLTIDROHNTIMVATQISWNPDTIIFRSTDDGATWTRIDWTSYPNRSLRYVLDISAE 360  
QY 407 PWLTGFGVQPNPPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 466

Db 361 PWLTGFGVQPNPPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
QY 467 APWKGLIETAVNDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSPVFTTGTSDY 526  
Db 421 APWKGLIETAVNDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSPVFTTGTSDY 480  
QY 527 AELNPSIIVRAGSFQSSQPNDRHVAFASTDDGKQWFGSEPGVTTGGTVAASADGSRFV 586  
Db 481 AELNPSIIVRAGSFQSSQPNDRHVAFASTDDGKQWFGSEPGVTTGGTVAASADGSRFV 540  
QY 587 WAPGDPQPVVAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALNGTFYRSTDDGVT 646  
Db 541 WAPGDPQPVVAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALNGTFYRSTDDGVT 600  
QY 647 QPVAAGLPSSGAVGVMFHVPKGGDLWLAASSGLYHSTNGSSWSAITGVSAVNVGFG 706  
Db 601 QPVAAGLPSSGAVGVMFHVPKGGDLWLAASSGLYHSTNGSSWSAITGVSAVNVGFG 660  
QY 707 KSAFGSSYPVAVVGTIGVGTGAYRSDDCGTTWVLINDDQHOYGNWQOAITGDHANLRV 766  
Db 661 KSAFGSSYPVAVVGTIGVGTGAYRSDDCGTTWVLINDDQHOYGNWQOAITGDHANLRV 720  
QY 767 YIGTNGRGIVYGDIGGAPSG 786  
Db 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 6  
US-10-155-400-6  
; Sequence 6, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
US-10-155-400-6

Query Match 78.6%; Score 4036; DB 14; Length 740;  
Best Local Similarity 100.0%; Pred. No. 6.9e-227;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 ATTPQYTWNSNVAIGGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWVGW 106  
Db 1 ATTPQYTWNSNVAIGGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWVGW 60  
QY 107 NNWYNGVWSIAADPINTNKWAAVGMVYNSWDPNDGAILRSSDQGTATWQITPLPFKLG 166  
Db 61 NNWYNGVWSIAADPINTNKWAAVGMVYNSWDPNDGAILRSSDQGTATWQITPLPFKLG 120  
QY 167 NMPGRGMLERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNPPDVGTYIANPTDT 226  
Db 121 NMPGRGMLERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNPPDVGTYIANPTDT 180  
QY 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNNPVFWSRSDGATWQAVPGATGF 286  
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNNPVFWSRSDGATWQAVPGATGF 240  
QY 287 IPHKGVDPVNVHLYIATSNTPGPDGSGDVWKFVSTGTWTRISPVSTDTANDYFGY 346  
Db 241 IPHKGVDPVNVHLYIATSNTPGPDGSGDVWKFVSTGTWTRISPVSTDTANDYFGY 300



Qy 347 SGLTIDRQHPNTIMWATQISWBDTIIIFRSTDCGATWTRIMWTSYPNRSIRYVLDISAE 406  
Db 301 SGLTIDRQHPNTIMWATQISWBDTIIIFRSTDCGATWTRIMWTSYPNRSIRYVLDISAE 360  
Qy 407 PMLTFGVQPNPPVPSPKLGMWDEMAIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 466  
Db 361 PMLTFGVQPNPPVPSPKLGMWDEMAIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
Qy 467 APVKGLIETAVNDLISPPSGAPLISALGDLGGTHADVTAVPSTIIFTSVPVFTGTSDVY 526  
Db 421 APVKGLIETAVNDLISPPSGAPLISALGDLGGTHADVTAVPSTIIFTSVPVFTGTSDVY 480  
Qy 527 AELNPSIIVRAGSFDPSQPNDRHVAFSTDGKNWFGSPPGGVTTGGTVAASADGSRFV 586  
Db 481 AELNPSIIVRAGSFDPSQPNDRHVAFSTDGKNWFGSPPGGVTTGGTVAASADGSRFV 540  
Qy 587 WAPGDCPQPVVYAVFGNSWAASQGVPPANAQIRSDRVPKTFYALSNGTYRSTDCGVTF 646  
Db 541 WAPGDCPQPVVYAVFGNSWAASQGVPPANAQIRSDRVPKTFYALSNGTYRSTDCGVTF 600  
Qy 647 QPVAAGLPSSGAVGVWVFHAPVPGKEDLWLAASGLYHSTNGSSWSAITGVSSAVNVGFG 706  
Db 601 QPVAAGLPSSGAVGVWVFHAPVPGKEDLWLAASGLYHSTNGSSWSAITGVSSAVNVGFG 660  
Qy 707 KSAPGSSYPVAVVGTIGVGTAYRSDDCGTTWVINDDQHQYGNWGOAITGDHANLRV 766  
Db 661 KSAPGSSYPVAVVGTIGVGTAYRSDDCGTTWVINDDQHQYGNWGOAITGDHANLRV 720  
Qy 767 YIGTNGRGIYVIGGAPSG 786  
Db 721 YIGTNGRGIYVIGGAPSG 740

## RESULT 7

US-10-156-761-9395  
; Sequence 9395, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 9395  
; LENGTH: 882  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-9395

Query Match 48.3%; Score 2478; DB 14; Length 882;  
Best Local Similarity 50.4%; Pred. No. 4.8e-136;  
Matches 473; Conservative 128; Mismatches 227; Indels 110; Gaps 14;  
Qy 23 TASFAVAAL--GVLPIATITASPAHAATTQPYTWSNVAGGGFVGDIVFNECAPGLTYV 80  
Db 8 TAVLALAAGLPACTPPALAAATATATADTYISWKNARVDGGFVGVIVNREKMLAYA 67  
Qy 81 RTDIGMYRWAAGRWPILLDVGWNNWNGYVIAADPINTKNVAAVGYMTNSWDP 140  
Db 68 RTDIGGAYRWAESSKTWPTLLDSVGWSDWCHTGTGVSLASDSDVPKNKYAAVGYTNSWDP 127

Qy 141 NDGAILRSSDOGATWQITPLPKLGNMPGRGMBERLAVDNNDNLIYFGAPSGKGLWS 200  
Db 128 GNGAVLRSRSDRGASQWKTDLFPKLGNNMPGRGMBERLAVDNPNRNVLYLFGAPSGKGLWS 187  
Qy 201 TDSGATWSQMTNFPDVGTVIANPTDTTCYQSDIQGVWVAFDPSKSSSSILGOAKTIFVGVVA 260  
Db 188 TDSGASWSQVTFPFPVGVYVQDADTSGYASDNQGIWVTFDESTGSGSSTRTVYGVVA 247  
Qy 261 DPNNPFWMSRDGGATWQAVPGAPTGFIPHKGVDFPNNHVLXIATSNITGTPYDGSDDVWK 320  
Db 248 DKNSVYRSTDAAGTWSRLAQPTGHLAKHGVLDAAAGCLYLAYSQKGGPYDGGKQLWR 307  
Qy 321 FSVTSGTWTRISPVSTDTANDYFGYSLTIDRQHPNTIMWATQISWBDTIIIFRSTDCG 380  
Db 308 YTTKTGTWNTISPAEADT---YTGFSGLTVDRQHPGTVMATAYSSWNPDTQLFRSTDSG 364  
Qy 381 ATWTRIDWTSYPNRSIRYVLDISAEPMWLTGCVQPNPPVPSPKLGMWDEMAIDPNSDR 440  
Db 365 GTWTKANDYTSYPSRSNRFTMDVSSPMLTWGANPAPPEQTPKLGMMTESLEIDPPDSAR 424  
Qy 441 MLYGTGATLYATNDLTKWDSGGQIHIAPMWKLEBETAVNDLISPPS--GAPLISALGDLGG 499  
Db 425 MMYGTGATVGTDLNLTWDSGSQFTIKPMARGLEBETAVNDLASPSGGQLFSALGDLGG 484  
Qy 500 FTHADVTAVPSTIIFTSVPVFTTGTSDVYAEALNPSIIVRAGSFDPSQPNDRHVAFSTDGK 559  
Db 485 FRHTDLITVPSLMYTSNFTTSTSLDYAETDPTGVVVRVGNLD--SGP---HVAFSTDNGA 539  
Qy 560 NWFQSGSPGGVTTGGTVAASADGSRFVWAPDGPQPVVYAVFGNSWAASQGVPPANAQIR 619  
Db 540 NWFAGADPSGVSGGTVAASADGSRFVWSPAGTG--VQYTTGFGTWSASAGLPAGAIVE 597  
Qy 620 SDRVNPKTFFALSNGTFYRSTDCGVTFQPVAA--GLPSSGAVGVWVFHAPVPGKEDLWLAAS 678  
Db 598 SDRVDPKTFYFKSGRFYVSSDGGATFTASAATGLPSGDS--VRKALPFTGKDIWLAGG 655  
Qy 679 S-----GLYHSTNGSSWSAITGVSSAVNVGFGSAPGSSYPVAVVGTIGVGTGAYRSD 733  
Db 656 ASDCAYGLWHSITDGAATFKLATVDQADTIGFKAATGASVQTLTYSAKIGGVGRIFRST 715  
Qy 734 DCGTTWVILINDDQHQYGNWGOAITGDHANLRVYIGTNGRGIYVIGDIGAPSGSPSPSVS 793  
Db 716 DKGASWTRVNDDAHQWGWGTGAATGDRPVYKVVYVSTNGRGIYVIGDTAGSSDCG----- 769  
Qy 794 PSASPSLSP 853  
Db 770 ----- 769  
Qy 854 SSPSSSP 913  
Db 770 ---GTEPAPT-----GACTVTTTITNMQSGGFQ--ADVQLANTGSTAWDQWSIG---W 814  
Qy 914 FTRDGGSTLVYNCDDWAAIGCG-----NIRASFSG 943  
Db 815 SFGDQGVETQLMNASYQAAGSGVTAANLAWNRVAAGS 852

## RESULT 8

US-09-917-376-7  
; Sequence 7, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL R.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1

```

; SEQ ID NO 7
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Aspergillus aculeatus
US-09-917-376-7

Query Match      32.7%; Score 1680; DB 12; Length 726;
Best Local Similarity 46.2%; Pred. No. 1.1e-89;
Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;

QY 47 ATTQPTWNSVAI--GGGGFVDGIVFNEGAGLIVRTDGGMYRWDAAANGRIPLLDWVG 105
DB 1 AASQAYTWKNVVTGGGGGFTPGIVFNPAGVAVARTDIGAYRLN--SDTWTPLMDWVG 59
QY 106 ---WNNNGYNGVSIADPINTNKVAAVGMVNTNSWDPNDGAILRSSDQAGATWQITPLPF 162
DB 60 NDTHWDM---GIDALATDPVDTDRVYVAVGMVNTNEWDPNVGSILRSIDQGTWETKLPF 116
QY 163 KLGNNPGRGMRGLAVDPNNNDNILYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTIYAN 222
DB 117 KVGNNPGRGMRGLAVDPNKNLSILYFGARSGHGLWKSIDYGATWSNVTSTFTWTGYFQD 176
QY 223 PTDITGYSQDIQGVVWVAFPKSSSLGQASKTIIFGVADPNPVPFWSRDCGATWQAVPGA 282
DB 177 SSST--YTSDPVGIWVTFDSTSGSGSATPRIFGVADAGKSVFKSEDAAGATWAVSGE 234
QY 283 PT-GFIPHKGVDPVNVHLYIATNTGPGYDGGSDVWKFVSCTGTWTRISPPVSTDTAN 341
DB 235 PQYGFPHKGVLSPEEKTLYISYANGAGPYDGTNGTVHKYNTISGVWTDISP---TSLAS 291
QY 342 DYFGYGLTIDRQHPNTIMVATQISWMPDTIIFRSTDCGATWTRIDWTSYPNRSRLRYVL 401
DB 292 TTYGYGGLSVLDQVPGTLWVAALNCWPDDELIFRSTDSGATWSPIWENWGYPSINYYSY 351
QY 402 DISAEPWLTGQVQPNP--PVSPKLGWMDMAIDPFNSDRMLYCTGATLYATNDLTKWDS 460
DB 352 DISNAPMIQDITSTDDQFPV---RVGMVVEALADPFDSNHLVYGLTLYVGGHDLTNWDS 408
QY 461 GGOIHIAPMVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTT 520
DB 409 KHNVTVKSIAVGIEMAVLGLITPPGFPALLSAVGDDGGFYHSDLDAAPNQAYHTPYGT 468
QY 521 GTSVDYAEINPSIIIVRAGSPDPSSQPNDRHVAFSTDGKKNMFOGSEPGVTTGGTVAASA 580
DB 469 TNGIDYAGNKPNSIVRSGASD--DYPT--LALSSNFGSTWYADYAASTGTGVALSA 523
QY 581 DGRFRWAPGDPQPVVAVVGFNGSNAASQGVPAQAQIRSDRVNPKTFYALNSGTFRST 640
DB 524 DGDVTLLMSSTSGALVSKSQG---TLTAVSSLPSGAVIASDKSDNTVYFGSAGAIYVSK 580
QY 641 DGGVTQPVAAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGSGSSWSAI--TGVSS 699
DB 581 NTATSFKTVS--LGSSTTVNAI--RAHPSIAGDVWASTDKGLMHSIDYGSTFTQIGSGVTA 638
QY 700 AVNVGFGKSAFGSSYPVAVVGTIGVGTGAYRSDDCGTTWVLINDDOHQYGNWQA--IT 757
DB 639 GWGFGFGKASSTGSVYVIYGFFTIDGAAGLKFSEDACTNMQVSDASHGFGS--GSANVNV 697
QY 758 GDHANLRVYGTNGRGIVYGDIGGAPSG 786
DB 698 GDLQTYGRVFRGHERPGHLLRQSQREPAG 726

```

## RESULT 9

```

US-10-155-400-7
; Sequence 7, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS

```

```

; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155.400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Aspergillus aculeatus
US-10-155-400-7

```

```

Query Match      32.7%; Score 1680; DB 14; Length 726;
Best Local Similarity 46.2%; Pred. No. 1.1e-89;
Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;

QY 47 ATTQPTWNSVAI--GGGGFVDGIVFNEGAGLIVRTDGGMYRWDAAANGRIPLLDWVG 105
DB 1 AASQAYTWKNVVTGGGGGFTPGIVFNPAGVAVARTDIGAYRLN--SDTWTPLMDWVG 59
QY 106 ---WNNNGYNGVSIADPINTNKVAAVGMVNTNSWDPNDGAILRSSDQAGATWQITPLPF 162
DB 60 NDTHWDM---GIDALATDPVDTDRVYVAVGMVNTNEWDPNVGSILRSIDQGTWETKLPF 116
QY 163 KLGNNPGRGMRGLAVDPNNNDNILYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTIYAN 222
DB 117 KVGNNPGRGMRGLAVDPNKNLSILYFGARSGHGLWKSIDYGATWSNVTSTFTWTGYFQD 176
QY 223 PTDITGYSQDIQGVVWVAFPKSSSLGQASKTIIFGVADPNPVPFWSRDCGATWQAVPGA 282
DB 177 SSST--YTSDPVGIWVTFDSTSGSGSATPRIFGVADAGKSVFKSEDAAGATWAVSGE 234
QY 283 PT-GFIPHKGVDPVNVHLYIATNTGPGYDGGSDVWKFVSCTGTWTRISPPVSTDTAN 341
DB 235 PQYGFPHKGVLSPEEKTLYISYANGAGPYDGTNGTVHKYNTISGVWTDISP---TSLAS 291
QY 342 DYFGYGLTIDRQHPNTIMVATQISWMPDTIIFRSTDCGATWTRIDWTSYPNRSRLRYVL 401
DB 292 TTYGYGGLSVLDQVPGTLWVAALNCWPDDELIFRSTDSGATWSPIWENWGYPSINYYSY 351
QY 402 DISAEPWLTGQVQPNP--PVSPKLGWMDMAIDPFNSDRMLYCTGATLYATNDLTKWDS 460
DB 352 DISNAPMIQDITSTDDQFPV---RVGMVVEALADPFDSNHLVYGLTLYVGGHDLTNWDS 408
QY 461 GGOIHIAPMVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTT 520
DB 409 KHNVTVKSIAVGIEMAVLGLITPPGFPALLSAVGDDGGFYHSDLDAAPNQAYHTPYGT 468
QY 521 GTSVDYAEINPSIIIVRAGSPDPSSQPNDRHVAFSTDGKKNMFOGSEPGVTTGGTVAASA 580
DB 469 TNGIDYAGNKPNSIVRSGASD--DYPT--LALSSNFGSTWYADYAASTGTGVALSA 523
QY 581 DGRFRWAPGDPQPVVAVVGFNGSNAASQGVPAQAQIRSDRVNPKTFYALNSGTFRST 640
DB 524 DGDVTLLMSSTSGALVSKSQG---TLTAVSSLPSGAVIASDKSDNTVYFGSAGAIYVSK 580
QY 641 DGGVTQPVAAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGSGSSWSAI--TGVSS 699
DB 581 NTATSFKTVS--LGSSTTVNAI--RAHPSIAGDVWASTDKGLMHSIDYGSTFTQIGSGVTA 638
QY 700 AVNVGFGKSAFGSSYPVAVVGTIGVGTGAYRSDDCGTTWVLINDDOHQYGNWQA--IT 757
DB 639 GWGFGFGKASSTGSVYVIYGFFTIDGAAGLKFSEDACTNMQVSDASHGFGS--GSANVNV 697
QY 758 GDHANLRVYGTNGRGIVYGDIGGAPSG 786
DB 698 GDLQTYGRVFRGHERPGHLLRQSQREPAG 726

```

## RESULT 10

```

US-10-420-191-2
; Sequence 2, Application US/10420191
; Publication No. US20040067569A1

```

GENERAL INFORMATION:  
; APPLICANT: No. US20040067569Allozymes Biotech, Inc.  
; APPLICANT: Rey, Michael W.  
; APPLICANT: Zaretsky, Elizabeth J.  
; APPLICANT: Haas, Jeffrey A.  
; TITLE OF INVENTION: Polypeptides Having Xyloglucanase Activity And Nucleic Acids  
; TITLE OF INVENTION: Encoding Same  
; FILE REFERENCE: 10210.200-US  
; CURRENT APPLICATION NUMBER: US/10/420,191  
; CURRENT FILING DATE: 2003-04-18  
; PRIOR APPLICATION NUMBER: US 60/373,987  
; PRIOR FILING DATE: 2002-04-19  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 2  
; LENGTH: 838  
; TYPE: PRT  
; ORGANISM: Trichoderma reesei  
US-10-420-191-2

Query Match 31.7%; Score 1625.5; DB 12; Length 838;  
Best Local Similarity 41.7%; Pred. No. 2e-86;  
Matches 350; Conservative 131; Mismatches 296; Indels 63; Gaps 23;

QY 25 SFAVAALGVLPITAITASPAHAATTPYTWNSVAI-GGGGFVDGIVFNEGAPGILYVRTD 83  
DB 4 SRVLALVIGAV-----IPAHAA-----FSWKNVKLGGGGFGVPGIIFHPKTKGVAVARTD 53

QY 84 IGMRYRDAANGRWIPLLDWV-----GNNNGYGVVSIADPINTNKVAAVGMVNTSWD 139  
DB 54 IGGLYRLN-ADDSWTAVTGDIADNAGHNNW---GIDAVALDPDQDKVYAAVGMVNTSWD 109

QY 140 PNDGAILRSSDCATQWITPLPKLGNMPPGRGWERLAVDPNNNDLILYFGASGGKGLWR 199  
DB 110 PSNGAILIRSDRGATWFSFTNLPPKVGNNMPGRGAGERLAVDPANSNIIFYFGASGNGLWK 169

QY 200 STDGATWMSQMTNFPDVGTYIANPTDTTQSDIQGVVWVAFKSSSSLSQAQSKTIFVGV 259  
DB 170 STDGGVTFKSVSFTATGTYIPDPSDNGYNSDKQLMWTFDSTSTTGATSRIFVGT 229

QY 260 ADP-NNPVFWSRDGGATQWAVPGAPGFIPIHKGVFDPVNHVLIATSNTPGPDGSGGDV 318  
DB 230 ADNITASVYVSTNAGSTWAVPGQPKYFPHKAKLQPAEKALYLTSDGTGPDGTLGSV 289

QY 319 WKFSVSGTWTRISPVSTDTANDYFCYSGLTIDRQHPNTIMWATQISWPPDIIIFRSTD 378  
DB 290 WRYDIAAGTWKDTITPVSGSDL---YFGFGLGLDQKPGTLVVASLNSWWPDQAFLPRSTD 346

QY 379 GGATWTRIMDWTSPNRSRLRYLDISAEPWLTFG-VQPNPPVPS---PKLGMWDEAMAI 433  
DB 347 SGTTWSPINAWASYPTEYYISSTPKAPWKNNFIDVTSESFDGLIKELGMWIESLEI 406

QY 434 DPFNSRMLYGTGATLYATNDLTWKDSGGQIHIAPMKVGLLEETAANDLISPPSGAPLISA 493  
DB 407 DPTDSNHLVYGTGWTIFGGHDLTNWDRHNSIQSLADGIEEFSVQDLASAPGSELLAA 466

QY 494 LGDLGGTHA---DVTAVPSTITFTSPTVFTGTSVDAELNPSIIVRAGSFPDSSQPNDRH 550  
DB 467 VGDNDGFTFASRNDLGTSPQWATPTWATSTSDYAGNSVKSVVRVGNAGTAGTQQ----- 521

QY 551 VAFSTDGKNWFGQSBEPGGVTTGTTVAASADGSRFVWAPGDPQPVVYAVFGFNSWAAQ 610  
DB 522 VALSSDGGATWSIDYAADTSMNGGTVAISADGDTILWSTASSG---VQSQFGSGFASVS 578

QY 611 GVPANAQIRSDRVNPKTFYALSNGTFYRSTGDTGTFQPVAAAGLPSSGAVGVW---FHAVPG 668  
DB 579 SLPAGAVIASDKKTNVSFFYAGSGTFVVKDGTGSSP---TRG-PKLSACTIRDIAAHTP 634

QY 669 KEGDLWLAASSGLYHSTNGGSSWAI-TGVSSAVNVGFGKAPGSSYPVAVFVGTGTGGVT 727  
DB 635 TAGTLYVSTDVGIFRSTDSGTTCQVSTALTNTYQIALGVGS-GSNW-NLYAFGT--GPS 690

QY 728 GA--YRSDDCGTTWVLINDDOHQYGNWQQAITGDHANLRVYIGTNGRGIVY--GDIGGA 783

DB 591 GARLYASGDSGASWTDIQSGQFGSIDSTKVAGSGSTAGQVTVGNRGRVFTAQGTGGG 750  
QY 784 PSGSPSPVSPSPASPSLSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 843  
DB 751 TGGT-----SSSTKQSSSTSSASSSTTLRSSVSVVTTTRASTVTSRTSSAAGPTGS 801

RESULT 11  
US-10-026-994-2  
; Sequence 2, Application US/10026994  
; Publication No. US20030113732A1  
; GENERAL INFORMATION:  
; APPLICANT: Dunn-Coleman, Nigel  
; APPLICANT: Goedegebuur, Frits  
; APPLICANT: Ward, Michael  
; APPLICANT: Yao, Jian  
; TITLE OF INVENTION: EGV1 Endoglucanase and Nucleic Acids  
; TITLE OF INVENTION: Encoding the Same  
; FILE REFERENCE: GC698  
; CURRENT APPLICATION NUMBER: US/10/026,994  
; CURRENT FILING DATE: 2002-04-30  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 818  
; TYPE: PRT  
; ORGANISM: Trichoderma reesei  
US-10-026-994-2

Query Match 31.2%; Score 1604; DB 14; Length 818;  
Best Local Similarity 41.8%; Pred. No. 3.5e-85;  
Matches 340; Conservative 128; Mismatches 291; Indels 54; Gaps 21;

QY 52 YTSNVAI-GGGGFVDGIVFNEGAPGILYVRTDTCGMTRWDAANGRWIPLLDWV---CW 106  
DB 2 FSWKNVKLGGGGFGVPGIIFHPKTKGVAYARTDIGLYRLN-ADDSWTAVTGDIADNAGW 60

QY 107 NNNCYNGVSVSIADPINTNKVAAVGMVNTSWDPNDGAILRSSDCATQWITPLPKLGG 166  
DB 61 HNW---GIDAVALDPDQDKVYAAVGMVNTSWDPNSGAILIRSDRGATWFSFTNLPPKVG 117

QY 167 NMPGRGAGERLAVDPNNNDLILYFGAPSGKGLWRSTDGATWMSQMTNFPDVGTYIANPTDT 226  
DB 118 NMPGRGAGERLAVDPANSNIIFYFGASGNGLWKSTDGVTGTSKVSFTATGTYIPDPSDS 177

QY 227 TGYQSDIQGVVWVAFDKSSSLQAQSKTIFVGVADP-NNPVFWSRDGGATQWAVPGAPG 285  
DB 178 NGVNSDKQGLMWVTFDSTSTTGGATSRIFVGTADNITASVYVSTNAGSTWAVPGQPK 237

QY 286 FIPHKGVFDPVNHVLIATSNTPGPDGSGGVKFSVTSGTWTRISVPVSTDTANDYFG 345  
DB 238 YFPHKAKLQPAEKALYLTSDGTGPTGLSVWRYDIAGGTWKDITPVSGSDL---YFG 294

QY 346 YSLGTIDRQHPNTIMWATQISWPPDITIFRGTGATWTRIMDWTSPNRSRLRYLDISA 405  
DB 295 FGLGLDLQKPGTLVVASLNSWWPDQAFLPRSTDGTTWSPINAWASYPTEYYISISTPK 354

QY 406 EPWLTFG-VQPNPPVPS---PKLGMWDEAMAI DPFNSRMLYGTGATLYATNDLTWKDS 460  
DB 355 APWKNNFIDVTSESFDGLIKELGMWIESLEIDPTDSNHLVYGTGWTIFGGHDLTNWDT 414

QY 461 GGQTHIAPMKVGLLEETAANDLISPPSGAPLISALDGLGPTH---DVTAVPSTITFTSPV 517  
DB 415 RHNVSIQSLADGIEEFSVQDLASAPGSELLAAVGDNDGFTFASRNDLGTSPQWATPT 474

QY 518 FTTCTSVYDAELNPSIIVRAGSFPDSSQPNDRHVAFTDGGKWFQGSSEPGVGTGGTVA 577  
DB 475 WATSTSDYAGNSVKSVVRVGN-----TAGTQVAISSDGGATWISIDYAADTSMNGGTVA 528

QY 578 ASADGRFVWAPGDPQPVVYAVFGFNSWAAQSGVPANAQIRSDRVNPKTFYALSNGTFY 637  
DB 529 YSADGDTILWSTASSG---VQSQFGSGFASVSSLPAGAVIASDKKTNVSFFYAGSGSTFY 585

QY 638 RSTDCGVTQPVAAGLPSSGAVGM--FHAVPCKEGDLMLAASSGLYHSTNGSSWSAI- 694  
DB 586 VSKDTGSSP--TRG-PKLSAGTIRIDIAAHPPTAGTLYVSTDVIGIFRSTDSGTTFGQVS 641  
QY 695 TCVSSAVNVGFKSAPGSSYPVAVVGTGGTGA--YESDDCGTTWVLINDDHOYGNW 752  
DB 642 TALTNTYQIALGVGS--GSNW-NLIYAPGT--GPSGARLYASGSGASWTDIQSGQGFSGID 697  
QY 753 GQAITGDHANLARVYGTNGRGIVY--GDIGAPSGSPSPSPSPSPSPSPSPSPSPSSP 810  
DB 698 STKVAGSGTAGVYGTNGRGVGYAQTGGTGT-----SSSTKQSSSSTSSA 748  
QY 811 SP 843  
DB 749 SSSSTLRSSWSTTRASTVTSSRTSSAAGPTGS 781

RESULT 12  
US-10-156-761-10111  
; Sequence 10111, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 10111  
; LENGTH: 739  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-10111

Query Match 28.1%; Score 1442; DB 14; Length 739;  
Best Local Similarity 40.8%; Pred. No. 8.6e-76;  
Matches 328; Conservative 107; Mismatches 272; Indels 96; Gaps 25;

QY 7 IRLTWRSRLVSLAATASFAVAALGVLPITATSPAHAAAT--QPYTWSNVAIGGGG 63  
DB 1 MRPSPSRRTV--LACTAA--AAALTAVP--AVGSOAHAAETTAGPSYRWRNAVIGTG 53  
QY 64 FVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWWGNNGYNGVVSIAADPIN 123  
DB 54 FVTGVLFPSPVGLAVARTDIGAYRWDGRTPLIDHLGWDNLLGVEAMVDPDTH 113  
QY 124 TNKWAACVMTNSWDPNGAILRSSDQATWQITPLPKLGNMPCRGMRGERLAVDPNN 183  
DB 114 PRLYLAVGTAAQSWAGN--GAVLRSEDRGATWTRTDLTKLGGNEDGRGAGERLLVDPRD 172  
QY 184 DNILYFGAPSGKLMRSTDSGATWSQMTNPPDVGTVIANPTDTTGYQSDIQGVVWVAFDK 243  
DB 173 STLWLGT--RHDGLLSTDRGATAAATAPP-----AKANSQQGVF----- 214  
QY 244 SSSSLGQAASKTIFVGVADPNPNP-----VFWMSRGGATWQAVPGAPTGF----IPHKGVDP 295  
DB 215 ----LVAAGRTVYAGWDGDTSGTANLYRTAD-GTTWGVPRPSGTSKAVPLRAYDT 269  
QY 296 VNHVLIATSNCTGGPYDGGSGDWKFSVSGTWTRISPV-----PSTDTANDYFGYGLTI 351  
DB 270 HTRELVTYTGADPGPGQSDGSHVHLKRTATGTWTEVTPVPGGTTSDGSADTFAYGGVAV 329

QY 352 DROHPNTIMVATQISWMPDTIIIPSTDGGATWTRIMDWTSYPNBSRLRYVLDISASPMLTF 411  
DB 330 DARRPGTLVSTNNRWADGDTVPRSTDGGRWTSLKO-----AAVFDVSETPFELDW 380  
QY 412 GVQPNPVPSPKLGWMDENMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPMVK 471  
DB 381 GDD-----KPKFGWQLQALAVDPYDSQHVYVYGTGATLYGTDLKRW-----APRIR 426  
QY 472 GLEETA VNDLISPPSG--APLISALGDLGGFTHADVAVPST-IFTSPTFTTGTSDVYABL 529  
DB 427 GLEESA VRQLISPPVGEAHLISGLDIGNVMYHERLTASPSRGMATNPVFGSATGLAQAAA 486  
QY 530 NPSIIVPAGSFDSSQPNDRHVAPESTDGKWM--FQSEPGGVTTGGTVAASADSRFW 587  
DB 487 RPAYVVRTGMDHGNG-----AYSHDGGRTWAPFEAQPDIAKADGAPGIATSDGGLTW 540  
QY 588 A----PGDPGPVVYA----VFGNSWAASQGVANAOIRSDRVNPKTFYA--LNGTFY 637  
DB 541 SFVHWDG-----TTYAAHRSTDNGASWSSEVSFPKATPVADPADPTFYAYDFDNGTLY 595  
QY 638 RSTDGVTQFQVAAAGLPSSGAVGMFHAVPKGEGDLMLAAS--SGLYHSTNGSSWSAITG 696  
DB 596 ASTDSGRSFTARAGLP--SGDSQFKLVAAAPCRSGDLMLSAKWNGLYRSTDDGDTTFARIDS 654  
QY 697 VSSAVNVGPKSAPGSSYPVAVVGTGGTTCAYRSDDCGTTWVLINDDHOYGNWQAI 756  
DB 655 CWASYTLGFGKAADGADYPAIYQVGTETITAVYRSDDAARTWVRINDDAHQWIGEA 714  
QY 757 TGDHANLRRVVYGTNGRGIVYGD 779  
DB 715 VGDPRHGRVYLTATNGRGIQYGE 737

RESULT 13  
US-10-395-241-12  
; Sequence 12, Application US/10395241  
; Publication No. US20040038367A1  
; GENERAL INFORMATION:  
; APPLICANT: YAOI, Katsuro  
; APPLICANT: MITSUISHI, Yasushi  
; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME  
; FILE REFERENCE: Q73756  
; CURRENT APPLICATION NUMBER: US/10/395,241  
; CURRENT FILING DATE: 2003-03-25  
; PRIOR APPLICATION NUMBER: JP 2002-83433  
; PRIOR FILING DATE: 2002-03-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 12  
; LENGTH: 812  
; TYPE: PRT  
; ORGANISM: Geotrichum sp. M128  
US-10-395-241-12

Query Match 21.9%; Score 1125; DB 12; Length 812;  
Best Local Similarity 34.5%; Pred. No. 2.8e-57;  
Matches 285; Conservative 129; Mismatches 291; Indels 120; Gaps 31;

QY 31 ALGVLPITATSPAHAAATQPYTWSNVAIGGGVFDGIVFNEGAPGILYVRTDIGMYRW 90  
DB 6 SLGKALTALSILASLAVAKEHYEFKNVAIGGGVITGIVAHPKTKDILLYARTDIGAYRW 65  
QY 91 DAANGRWIPLLDWWGNNGYNGVVSIAADPINTKWAAVCMYT--NSWDPNDGAILRSS 149  
DB 66 DAGTSKWIPLDLFTEAQDMNIMGTSTALDPNNDRLYLAQGRYVGDEW----AAFYVSE 121  
QY 150 DQATWQITPLPKLGNMPCRGMRGERLAVDPNNNDNLYFGAPSGKLMRSTDSGATWSQ 209  
DB 122 DRGQSFYIESPPFGANDMGRRNGERLAVNPFFNSVWMTGRT--EGIKWSSDRAKTWN 180  
QY 210 MTFNPDPVGTVIANPTDTTGYQSDIQGVVWVAFDSSSSSLGQAASKTIFVGVADPNPNPVFS 269

Db 181 VTSIPDAF-----TNGIGYTS-----VIFDP-----ERNGTIYASATAPQG-MYVT 220  
Qy 270 RDGATQWAVPGAPTGI-----PHKGVDPVNVHLYIATSNTPGPY 311  
Db 221 HDGGVSWEPVAGQPSWLNRTTGAPDKKASIAQPMKVALTP--NFLVTVADYPCPW 278  
Qy 312 DGSAGVWKFESVTSCTWTRI-----SPVSTDTANDYFGYSGLTITDRQHPNTIMVATQ 364  
Db 279 GVTFGEVWRQNRITSGAWDDITPRVGNSSPAPYNNQTPPAGGFCGLSDATNPRLVVIT- 337  
Qy 365 ISWMPDTI---IFRSTDDGATWRIWDWTS-----YPNRSRVLVDISAEPMLTFG 412  
Db 338 LDRDPGALDSIYLSIDAGATWQDVQLSSPSNLEGNWGHPTNAARY-KDGTVPVWLDNF 396  
Qy 413 VQP-----NPPVPSP---KLGMNDEAMADFPNSDRMLYGTGATLYATNDLTK-----WDS 460  
Db 397 NGPQWGGYGAPHCTPGLTKEFGWMSAVLIDFPNPEHLMYGTGATIMATDILSRVEKDM-- 454  
Qy 461 GGOIHIAP---MVKGLEETAVNDLISPPSGAPLISALGDLGTFHADTVAVPSTIPTSP 516  
Db 455 -----APSWYLQIDGIEENAILSLRSPKSGAALLSGIDISGMKHDLLTK-PQKMEGAP 507  
Qy 517 VFTTGTSDVYAEINPSIIVRAGSFPDSSQPNDRHVAESTDGGKNW--FQSGEPGGVTT-- 572  
Db 508 QFNLDSIDAAGNFPNVPVVRAGSSGHEYDSACARGAYATDGGDAWTFPTCPGMNASHY 567  
Qy 573 -GCTVAASADGSRFVWAP--GDQPQPVYAVGFCNSWAASQGVPA-----NAQIRSDRV 623  
Db 568 QGSTIADVAGSGQIVWSTKLDEQASGPWYSHDYGTWS-----VPAGDLKAQATANVLSDKV 623  
Qy 624 NPRTFYALNGTFTYRSTDDGVTQPQVAAGLPSSGAVGMFHAVPGKBDLWL-AASSGLY 682  
Db 624 QDCTFYATDGGKFFVSTDDGKSYAAKAGGLVT--GTSLMPAVNVPWVAGDVVVPVEGGLF 681  
Qy 683 HSTNGSSWSAI-TGVSSAVNVGFGKS-----APGSSYPVAVFVGT--IGVTVGAYRSDDC 735  
Db 682 HSTDFGASFTVGTANATLVSVGAPKSKDGKASAPSAVFIWGTDKPGSDIGLYRSDDN 741  
Qy 736 GTTWVLINDHOYGNWQQAITGDHANLRVYIGTNGRGIVYGD 780  
Db 742 GSTWTRVNDQEHNSYSG-PTWIEADPKVYGRVYLTNGRGIVYADL 785

## RESULT 14

US-10-395-241-14  
; Sequence 14, Application US/10395241  
; Publication No. US20040038367A1  
; GENERAL INFORMATION:  
; APPLICANT: YAOI, Katsuro  
; APPLICANT: MITSUISHI, Yasushi  
; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
; FILE REFERENCE: Q73756  
; CURRENT APPLICATION NUMBER: US/10/395,241  
; PRIOR FILING DATE: 2003-03-25  
; PRIOR FILING DATE: 2002-03-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 14  
; LENGTH: 789  
; TYPE: PRT  
; ORGANISM: Geotrichum sp. M128  
US-10-395-241-14

Query Match 21.8%; Score 1117; DB 12; Length 789;  
Best Local Similarity 35.0%; Pred. No. 7.9e-57;  
Matches 281; Conservative 124; Mismatches 279; Indels 120; Gaps 31;  
Qy 52 YTHSNVAICGGFVDGIVFNEGAPGILYVTRTDIGMYRWDAAANGRWIPLDWDYGNWGY 111  
Db 4 YEPKNAIGGGVITGIVAHPKTKLLYARTDIGGAYRWDAGTSKWIPLNDFFEAQDMNI 63

Qy 112 NGVVSIAADPINTNKVAAVCMYT-NSWDPNDGAILRSSDQATWQITPLPKLGNMMPG 170  
Db 64 MGTESIALDPNPNRLYLAAQRYGVGDEW-----AAFVVSSEDRCQSFYIYSSPFPMGANDMG 119  
Qy 171 RGMERLAVDNNNDILYFGAPSGKGLWRSTDSGATWSOMTNFDPDVGTVIANPTDITGYQ 230  
Db 120 RANGERLAVNPNSEVWVGTRT-EGIKSSDRAKTWTNVTISIPDAF-----TNGIGYT 172  
Qy 231 SDIQGVVWVAFDKSSSLGQASKITFVGVADPNPNFVMSRDGGATWQAQVPGAPTGI--- 287  
Db 173 S-----VIFDP-----ERNGTIYASATAPQG-MYVTHDGGVSWEPVAGQPSWLNRT 218  
Qy 288 -----PHKGVDPVNVHLYIATSNTPGPGYSGSDGVKWSVTSCTWTRI- 331  
Db 219 TGAPFDKPKASIAQPMKVALTP--NFLVTVADYFGPVGVTGFEVWRQNRITSGAWDDIT 276  
Qy 332 -----SPVSTDTANDYFGYSGLTIDRQHPNTIMVATQISWMPDTI---IFRSTDDGAT 382  
Db 277 PRVGNSSPAPYNNQTPPAGGFCGLSDATNPRLVVIT-LDRDPGALDSIYLSIDAGAT 335  
Qy 383 WTRIWDWTS-----YPNRSRVLVDISAEPMLTFGVQP-----NPPVPSP---KL 425  
Db 336 WKDVTQLSSPSNLEGNWGHPTNAARY-KDGTVPVWLDNFNGPQWGGYGAPHCTPGLTKEG 394  
Qy 426 WMDAMADFPNSDRMLYGTGATLYATNDLTK---WDSGGQIHIAP-----MVKGLEETA 477  
Db 395 WMSAVLIDFPNPEHLMYGTGATIMATDILSRVEKDM-----APSWYLQIDGIEENA 446  
Qy 478 VNDLISPPSGAPLISALGDLGTFHADTVAVPSTIFTSPVFTTGTSDVYAEINPSIIVRA 537  
Db 447 ILSLRSPKSGAALLSGIDISGMKHDLLTK-PQKMEGAPQFNLDSIDAAGNFPNVPVRA 505  
Qy 538 GSPDPSQPNDRHVAESTDGGKNW--FQSGEPGGVTT--GCTVAASADGSRFVWAP--G 590  
Db 506 GSSGHEYDSACARGAYATDGGDAWTFPTCPGMNASHYQGSTIADVAGSGQIVWSTKLD 565  
Qy 591 DPGQPVYVAVGFCNSWAASQGVPA-----NAQIRSDRVNPKTFYALNGTFTYRSTDDGV 644  
Db 566 EQASGPWYSHDYGTWS-----VPAGDLKAQATANVLSDKVQDCTFYATDGGKFFVSTDDGK 621  
Qy 645 TFQPVAAGLPSSGAVGMFHAVPGKBDLWL-AASSGLYHSTNGSSWSAI-TGVSSAVN 702  
Db 622 SYAAKAGGLVT--GTSLMPAVNVPWVAGDVVVPVEGGLFHSSTDFGASFTRVGTANATLVS 679  
Qy 703 VGFGKS-----APGSSYPVAVFVGT--IGVTVGAYRSDDC GTTWVLINDHOYGNWQQA 756  
Db 680 VGAPKSKDGKASAPSAVFIWGTDKPGSDIGLYRSDDN GSTWTRVNDQEHNSYSG-PTWI 738  
Qy 757 TGDHANLRVYIGTNGRGIVYGD 780  
Db 739 EADPKVYGRVYLTNGRGIVYADL 762

## RESULT 15

US-10-395-241-18  
; Sequence 18, Application US/10395241  
; Publication No. US20040038367A1  
; GENERAL INFORMATION:  
; APPLICANT: YAOI, Katsuro  
; APPLICANT: MITSUISHI, Yasushi  
; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
; FILE REFERENCE: Q73756  
; CURRENT APPLICATION NUMBER: US/10/395,241  
; CURRENT FILING DATE: 2003-03-25  
; PRIOR APPLICATION NUMBER: JP 2002-83433  
; PRIOR FILING DATE: 2002-03-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 18  
; LENGTH: 826  
; TYPE: PRT  
; ORGANISM: Artificial Sequence



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 14, 2004, 08:53:04 ; Search time 9066 Seconds  
(without alignments)  
4575.257 Million cell updates/sec

Title: US-09-917-376-1  
Perfect score: 5135  
Sequence: 1 MDRSENIRLTMRSRLVSL.....RASFGSVNPATPTADTYLOX 957

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09917376/runat\_14052004\_085256\_4619/app\_query.fasta.1.1095  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=oct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09917376.@CGN\_1\_1\_5932@runat\_14052004\_085256\_4619 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -XGAPOP=10 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sta:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_ox:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sta:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	5134	100.0	2869	6	AX700058	AX700058 Sequence
2	2493	48.5	296500	1	SC039128	AL939128 Streptomy
3	2487	48.4	299800	1	AP005028	AP005028 Streptomy
4	2419.5	47.1	3262	1	AF078038S1	AF078038 Caldicell
5	2174	42.3	2823	6	AX565635	AX565635 Sequence
6	2037	39.7	12732	1	AE007608	AE007608 Clostridi
7	2007.5	39.1	2950	1	AJ585344	AJ585344 Clostridi
8	1784	34.7	2579	8	AB1292929	AJ292929 Agaricus
9	1782.5	34.7	2745	8	AB015511	AB015511 Aspergill
10	1690	32.9	3959	8	AY040839	AY040839 Aspergill
11	1633	31.8	2724	8	AK110506	AK110506 Oryza sat
12	1631.5	31.8	2517	8	AY281371	AY281371 Hypocrea
13	1570	30.6	299550	1	AP005031	AP005031 Streptomy
14	1179.5	23.0	14520	1	AE011809	AE011809 Xanthomon
15	1170.5	22.8	11910	1	AE012276	AE012276 Xanthomon
16	1161	22.6	13278	1	AE001712	AE001712 Thermotog
17	1125.5	21.9	2646	6	AX924429	AX924429 Sequence
18	1125.5	21.9	2716	8	AB089343	AB089343 Geotrichu
19	1124	21.9	2481	6	AX924435	AX924435 Sequence
20	1117	21.8	2367	6	AX924431	AX924431 Sequence
21	823	16.0	3687	6	AX700036	AX700036 Sequence
22	702	13.7	2289	6	AX700050	AX700050 Sequence
23	460.5	9.0	590	8	AB1534359	AB1534359 Agaricus
24	431.5	8.4	5437	1	CASR69XYN2	AF036924 Caldicell
25	429	8.4	3365	6	AX700025	AX700025 Sequence
26	409	8.0	11707	1	AF078737	AF078737 Caldicell
27	409	8.0	11707	6	E35099	E35099 Truncated c
28	407	7.9	299175	1	AP005023	AP005023 Streptomy
29	402.5	7.8	3237	1	AF200304	AF200304 Caldicell
30	398.5	7.8	5513	1	AT286105	Z86105 A.thermophi
31	397.5	7.7	5439	1	DDCCBLA	L32742 Caldocellum
32	391.5	7.6	5284	1	CSU16308	UI6308 Caldocellum
33	389	7.6	24208	1	AY281357S2	AY281358 Unculture
34	387	7.5	344615	1	EX569695	EX569695 Synchoco
35	381.5	7.4	2801	5	AF218784	AF218784 Gallus ga
36	380	7.4	209091	10	AC122229	AC122299 Mus muscu
37	372.5	7.3	2015	8	CREVSP3A	L29029 Chlamydomon
38	370	7.2	206849	2	AC011760	AC011760 Drosophil
39	367.5	7.2	215910	10	AC127335	AC127335 Mus muscu
40	367	7.1	7085	2	AC015260	AC015260 Drosophil
41	367	7.1	179386	3	AC104515	AC104515 Drosophil
42	367	7.1	273948	3	AE003574	AE003574 Drosophil
43	365	7.1	3768	2	AC015261	AC015261 Drosophil
44	365	7.1	117167	9	AC110813	AC110813 Homo sapi
45	365	7.1	240904	2	AC126204	AC126204 Rattus no

ALIGNMENTS

RESULT 1



AX700058  
LOCUS AX700058 2869 bp DNA linear PAT 03-APR-2003  
DEFINITION Sequence 2 from Patent WO03012090.  
ACCESSION AX700058  
VERSION AX700058.1 GI:29536021  
KEYWORDS  
SOURCE Acidotherrmus cellulolyticus  
ORGANISM Acidotherrmus cellulolyticus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Frankineae; Acidotherrmaceae; Acidotherrmus.  
REFERENCE  
1 Ding, S.Y., Adney, W.S., Vinzant, T.B. and Himmel, M.E.  
Thermal tolerant avicelase from *Acidotherrmus cellulolyticus*  
Patent: WO 03012090-A 2 13-FEB-2003;  
Midwest Research Institute (US)  
FEATURES  
Location/Qualifiers  
source 1..2869  
/organism="Acidotherrmus cellulolyticus"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:28049"  
ORIGIN  
Alignment Scores:  
Pred. No.: 3,44e-147 Length: 2869  
Score: 5134.00 Matches: 956  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.98% Indels: 0  
DB: 6 Gaps: 0  
US-09-917-376-1 (1-957) x AX700058 (1-2869)  
QY 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgLeuValSerLeuLeu 20  
DB 1 ATGGATCGTTCGGAGAACATCCGCTGCTGACTATGAGATCAGACATGGGTACTCTC 60  
QY 21 AlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThr 40  
DB 61 GCGCCACTCGCTGCTTCGCGCGGCGCGCTCTGCGAGTCTGCGCCATCGGATAACG 120  
QY 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGly 60  
DB 121 GCTTCCTCGCGACGCGCGGACGACTCAGCCGTACACCTGGAGCAACGTTGGCGATCGG 180  
QY 61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuVal 80  
DB 181 GCGCGCGCTTTGTCACGGGATCGCTTCAATGAAGTGCACCGGAATCTGTACGTG 240  
QY 81 ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu 100  
DB 241 CGGACGGACATCGGGGGGATGATCGATGGATGCGCCCAACGGCGGTGGATCCCTCTT 300  
QY 101 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp 120  
DB 301 CTGGATTGGGTGGATGGAAATTCGGGGTCAACCGCGCTCGTCAAGATTGGCGGACAG 360  
QY 121 ProIleAsnThrAsnIleValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPro 140  
DB 361 CCGATCAATCACTAACAGGTATGGCGCGCGCTCGGAATGTACCAACAGCTGGGACCCA 420  
QY 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu 160  
DB 421 AACGACGGAGCGATTTCCGCTGCTGTATCAGGCGCGCAACGTGGCAATAACGCCCTG 480  
QY 161 ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180  
DB 481 CCGTTCAGGCTTGGCGGCAACATCGCGCGGTGGAAATGGCGAGCGGCTTGGGTGGAT 540  
QY 181 ProAsnAsnAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSer 200  
DB 541 CCAAAATGACAAACATTTCTGTATTTTCGCGCGCGCGCGGCAAGGGCTCTGGAGAAGC 600  
QY 201 ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle 220

DB 601 ACAGATTCCGCGCGGACCTGTCCTCAGATGACGAACCTTTCGCGAGTGGACGACGATCAT 660  
QY 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla 240  
DB 561 GCAATCCCACTGACACGCGCTATCAGACCGATATTCAGGGGTCTGCTGGGTGGCT 720  
QY 241 PheAspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAla 260  
DB 721 TTCGACAACTCTTCGTCTCATCGCTCGGCGAAGCGAGTAAGCAATTTTGTGGCGGTGGCG 780  
QY 261 AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValPro 280  
DB 781 GATCCCAATATATCCGCTTCTTGAGCAGACAGACCGCGCGCGAGCTGGCAGCGGTGGCG 840  
QY 281 GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu 300  
DB 841 GGTGCGCGCGAGCTTCATCCCGACAGAGGCGCTTTTGACCGGTCAACACGCTGCTC 900  
QY 301 TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLys 320  
DB 901 TATATTGCCACGACCAATACGGGTGGTCCGATGACGGGAGCTCCGCGGACGCTCTGAAA 960  
QY 321 PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla 340  
DB 961 TTCTCGGTGACCTCCGGGACATGAGCGGAATCAGCCGCTACCTTCGACGACGACGCGC 1020  
QY 341 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360  
DB 1021 AACGACTACTTTGGTTACAGCGGCTCTACTATGACCGCCAGCACCCGAAACAGATAATG 1080  
QY 361 ValAlaThrGlnIleSerTrpTrpAspThrIleIlePheArgSerThrAspGlyGly 380  
DB 1081 GTGCAACCCAGATATCGTGTGCGCGGACACCATATCTTTTCGAGACCCGACGCGCT 1140  
QY 381 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 400  
DB 1141 GCGACGTGGACGCGGATCTGGGATTTGAGAGATTCCTCAATCGAAGCTTGGATATGTG 1200  
QY 401 LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro 420  
DB 1201 CTTGACATTTCCGCGGAGCTTGGCTGACCTTCGCGTACAGCCGAATCTCCCGTATCCC 1260  
QY 421 SerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 440  
DB 1261 AGTCCGAAGCTCGCTGGATGGATGAGCGATGCAATCGATCCGTTCACACTCTGATCG 1320  
QY 441 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 460  
DB 1321 ATGCTCTACGGAACAGCGCGACGTTGTACGCAACAAATGATCTCACGAAGTGGGACTCC 1380  
QY 461 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAsp 480  
DB 1381 GCGCGCGAGATTCATATTCGCGCCGATGGTCAAGGATTGGAGGAGACGCGGTAAACGAT 1440  
QY 481 LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 500  
DB 1441 CTATACGCGCGCTGTCGCGCGCCGCTCATCAGCGCTCTCGAGAGACTTCGCGCGCTTC 1500  
QY 501 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 520  
DB 1501 ACCCACGCGGACGTTACTGCGCTGCCATCGACGATCTTTCAGTCCCGGTGTTCAGACC 1560  
QY 521 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 540  
DB 1561 GGCACGACGCTCGACTATGCGGAATGGAATCCGTCGATCATCGTTCGCGCTCGGAAGTTTC 1620  
QY 541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 560  
DB 1621 GATCCATCGAGCCCAACCGAACGACGAGCGGACGCTCGGCTCTCGACAGACGCGCGCAAGC 1680  
QY 561 TrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAlaSerAla 580

Db 1681 TGGTTCAAGCGACGAACTCGCGGGGTGACGAGCGGGCGACCGTCGCGGCATCGGCC 1740

Qy 581 AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVal 600

Db 1741 GACGGCTCTGTTGCTGGGCTCCCGCGATCCCGTCAAGCTGTGGTGTAGCGAGTC 1800

Qy 601 GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAlaGlnIleArgSer 620

Db 1801 GGATTTGGCAACTCTCGGGTGTCTCGCAAGGTGTTCGCCCAATGCCAGATCCGCTCA 1860

Qy 621 AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr 640

Db 1861 GACCGGTGAATCCAAAGACTTTCTATGCTCTATCCCAATGAACTTCTATCAAGCAGC 1920

Qy 641 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 660

Db 1921 GACGGGGGTGACATTCNACCGGTGCGGGCGGTCTTCGAGCAGCGGTGCGTGGT 1980

Qy 661 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaSerSerGly 680

Db 1981 GTCATGTTCCAGCGGTGCTCGAAAAGAGCGATCTGTGGTCTGCTGCATCGAGCGG 2040

Qy 681 LeuTyrHisSerThrAsnGlySerSerTrpSerAlaIleThrGlyValSerSerAla 700

Db 2041 CTTTACCACCTCAACCAATGCGCGCAGCTGTGCTGCAATCAACCGCGTATCTCCGCG 2100

Qy 701 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 720

Db 2101 GTGAACGTGGGATTTGGTAAGTCTGCGCGCGGTGCTCATACCGAGCGGTCTTGTGCTC 2160

Qy 721 GlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTrpVal 740

Db 2161 GGCACGATCGAGCGGTACGGGGGCGTACCGTCCGACGACTGTGGGACGACTGGGTA 2220

Qy 741 LeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHis 760

Db 2221 CTGATCAATGATGACGACCAACCAATACGGAATTTGGGACACGAATCAACCGGTGACCA 2280

Qy 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780

Db 2281 GCGAATTTACGGCGGTGTACATAGGACACGAAACGCGGTGGAATGTATACGGGGACATT 2340

Qy 781 GlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeu 800

Db 2341 GGTGGTGGCGGTCGCGATCGCGCTCTCGTGGTGGTCTCGTGGCTTCGCGAGCGCTG 2400

Qy 801 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820

Db 2401 AGCCCGAGCCGAGCCGAGCAGCTGCCATCGCGTCCGCGTCCGCGAGCTCGAGTCCA 2460

Qy 821 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840

Db 2461 TCTCTGTCGCGGTCTCGTCCGCTCACCATCCGCGAGTCCGCTCGCTCCGCTCACCA 2520

Qy 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860

Db 2521 TCGCGCTCCCGAGCCGCTCTTCTGTCAGGCGCGTCTTCTGTCACCGTCTTCTGCGCG 2580

Qy 861 SerProThrProSerSerProValSerGlyGlyValLysValGlnTyrLysAsnAsn 880

Db 2581 AGCCCAACCGCTGCTGTCGCGCGGTGTCCGGTGGGGTGAAGTGCAGTATAAGATAAT 2640

Qy 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900

Db 2641 GATTCGGCGCGGGTGATTAATCAGATCAAGCCGGGTTTGCAGGTGTGAATACCGGGTCG 2700

Qy 901 SerSerValAspLeuSerThrValThrValTyrTrpPheThrArgAspGlyGlySer 920

Db 2701 TCGTGGGTGGATTTGTCGACGGTGACGGTGGGTACTGTGGTTTCAACCGGGATGTGGCTCG 2760

Qy 921 SerThrLeuValTyrAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSer 940

Db 2761 TCACACTGTTGTACAACTGTGACTGGCGCGCATCGGGTGTGGATATATCCCGCGCTCG 2820

Qy 941 PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 956

Db 2821 TTCGGCTCGGTGAACCCGCGACGCGCGGACACCTACCTCGCAG 2868

RESULT 2

SC0939128/c 296500 bp DNA linear BCT 11-FEB-2003

LOCUS Streptomyces coelicolor A3(2) complete genome; segment 25/29.

DEFINITION AL039128 AL032862 AL031013 AL031107 AL031350 AL031371 AL031515

ACCESSION AL033505 AL034443 AL035161 AL079355 AL096825 AL645882

VERSION AL939128.1 GI:24413907

KEYWORDS Streptomyces coelicolor A3(2)

SOURCE Streptomyces coelicolor A3(2)

ORGANISM Bacteria: Actinobacteria: Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S., Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrall, B.G., Parkhill, J. and Hopwood, D.A. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2) Nature 417 (6885), 141-147 (2002)

JOURNAL MEDLINE 21996410

PUBMED 12000953

REFERENCE 2 (bases 1 to 296500)

AUTHORS Bentley, S.D.

TITLE Direct Submission

JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk

COMMENT On or before Oct 26 2002 this sequence version replaced 91:20520750, 91:20520810, 91:20520753, 91:20520666, 91:20520757, 91:20520758, 91:20520668, 91:20520759, 91:20520814, 91:20520765, 91:20520767.

FEATURES

source 1. 296500

/organism="Streptomyces coelicolor A3(2)"

/mol\_type="genomic DNA"

/strain="A3(2)"

/db\_xref="taxon:100226"

complement (144..443)

/gene="SCO6480"

/note="synonym: SC9C7.16c"

complement (144..443)

/gene="SCO6480"

/note="SC9C7.16c, unknown, len: 99 aa"

/codon\_start=1

/translation="hypothetical protein SC9C7.16c"

/protein\_id="CAA22728.1"

/db\_xref="GI:4154075"

/db\_xref="SPTREMBL:Q92B04"

/translation="MTVGCIAASMTRPGFRERVSILLISAADGLRDAGCELYVVGISE DDETTIWTVEWTRERDHDASLELPAAKDAIGRAMPMLTGEFTKQEMSVAAGGLG"

complement (517..1563)

/gene="SCO6481"

/note="synonym: SC9C7.17c"

complement (517..1563)

/gene="SCO6481"

/note="SC9C7.17c, conserved hypothetical protein, len: 348aa; similar to several hypothetical proteins eg. SW:Y0BN MYCTU from Mycobacterium tuberculosis (340 aa) fasta scores; opt: 998, z-score: 1321.8, E(): 0, (47.6% identity in 355 aa overlap)."

/codon\_start=1

```

/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAA22729.1"
/db_xref="GI:4154076"
/db_xref="SPTREMBL:Q9ZBJ3"
/translation="MRWENLTSDHARAADTALFGADAVVTRTETDPERGITFEHVR
ARSILNRVPGASRPPEFTWVNRGCTHACVYCFARKTHSYLDLDTGLGDSQIVVVK
NAPEVLRLQASRWGQEVAMGTNVCYORAEGRVRLMFGPIIEALTERANPFSILTK
GTLILDLRLTRAAVTVSGISVSGFTDABLWRTVFPPTAPERFVRLALGERG
ICGGVLMAPVIFPLGDEPAQLRAIAAAGATSVTPLVHLRPGAREFWFMAWLGRRH
HPLHVRXYERLYADGAYAPKWQRRITRQVHDLAREYIGIPARAGMPRRVAPEHSRP
AEPAPPEPQLSLI"
1800..2243
/gene="SCO6482"
/notes="synonym: SC9C7.18"
1800..2243
/gene="SCO6482"
/notes="SC9C7.18, conserved hypothetical protein, len: 348
aa; similar to two hypothetical proteins e.g. TR:069685
(EMBL:AL022121) Mycobacterium tuberculosis H37Rv18c or
MTV025.066c, 147 aa; fasta scores; opt: 475, z-score:
768.3, E(): 0, (52.1% identity in 146 aa overlap)"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAA22730.1"
/db_xref="GI:4154077"
/db_xref="SPTREMBL:Q9ZBJ2"
/translation="MAQVETATRIVAADAETVDFDLADYSGTRELKLPHEHSEYEV
EGDGEGLVHWKLAITSVRKRDCLLETFEDTDELVDKDRNSWTTWRVTPAGEGK
SRVVLVTWQAGGVGGFFETKFPAPKGLRIYDALLAKLAEEVK"
2394..4841
/gene="SCO6483"
/notes="synonym: SC9C7.19"
2394..4841
/gene="SCO6483"
/notes="SC9C7.19, probable efflux protein, len: 815aa;
similar to many, several of which are involved in toxin
/antibiotic efflux eg. TR:069070 (EMBL:AF061335)
oxytetracycline exporter from Streptomyces rimosus (536
aa) fasta scores; opt: 1020, z-score: 879.9, E(): 0,
(36.9% identity in 542 aa overlap)"
/codon_start=1
/transl_table=11
/product="putative efflux protein"
/protein_id="CAA22731.1"
/db_xref="GI:4154078"
/db_xref="SPTREMBL:Q9ZBJ1"
/translation="MGGITLVDPVDPVQDEPVAQAQPPGPPVPDPGDERDAGLSP
RRVRLVFGMLALLALEOMIVATLPKIVGELHGLDKMSWATYALLTSTVGLPI
YKGLDLGCKGVQFQAILVYFVVSALAGRAQTMOLIAFRAVQGVGAGLMTGVQAI
IADVLPERRRRTMGLIGASFGLASVAGPELLGGYFTDHLNWRWCFFYINPFGILTVAV
IATVLPKPRVRKPRDILGALLAAASTCLVLLTWSGGTEYAWGSRVILGLAAGVA
ATTLFLAAEHPAPELPIRLFRDTSFNVTALVGLVGVALGASLYLFTLQMVDGA
SATESGLLPMWGVGASIIISGLISRTGHIYRHPILGSLSVVWMLLSRLDADT
SRLHSIMAWLIGAGLVMPVLAVQNSVRPTDGTATSNANYPROIGSGVGAFF
GTLFAGRLDALADRIAPAEANGVGLPAEALITQLVHSLPDLADYIGAYADAMPRI
LYLVPVLGLLFLKPKLVSHNAVPTDDETPAHQPQPAQVPAAPSPARSPYAGGI
PVCVTQHPDGTVPVPAALTLIDVGGQIGRGASGDDGRYALATPGSGAYVLIAAAG
HQPQAVSVTVGRPVLDVLVLAGAGLAGSVMTADGSPVRDADTTLTNVHGVEATTR
SGREGVITLTVAGYVTLAAAPAPRPAALPVSVQAAETRODVLAGGAVLRGTVR
ASGGRVEDARVTLTLDAGNVVDTLTGPDGFRFRVLDLSSGEYTVIAAGYPPVATVLQ
VAGGGRTERDLQLGHED"
4911..7493
/gene="SCO6484"
/notes="synonym: SC9C7.20"
4911..7493
/gene="SCO6484"
/notes="SC9C7.20, conserved hypothetical protein, len:
860aa; similar to many hypothetical proteins egs.
TR:068875 (EMBL:U22894) from Streptomyces lividans (547
aa) fasta scores; opt: 1000, z-score: 738.2, E(): 0,

```

```

(40.0% identity in 497 aa overlap) and TR:086807
(EMBL:AL031031) from Streptomyces coelicolor (916 aa)
fasta scores; opt: 358, z-score: 316.9, E(): 2.3e-10,
(27.2% identity in 779 aa overlap). Also contains a short
region of similarity to TR:Q53897 (EMBL:X60316) Ahaa
regulatory locus from Streptomyces coelicolor (192 aa)
fasta scores; opt: 108, z-score: 223.2, E(): 3.8e-05,
(38.5% identity in 135 aa overlap)".
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAA22732.1"
/db_xref="GI:4154079"
/db_xref="GOA:Q9ZBJ0"
/db_xref="SPTREMBL:Q9ZBJ0"
/translation="MVTGHPRTVVVTAQILRSRGQRGPDAMDHGTEDAPPGHAGPDP
TVPADPVTGRVPLAVVDRDGLVSHWSRGARRLFSKEAIGRPAPDLLLPVSGVLP
EDAEPPGAGHDDSYRSYTAHDGPGFGLSSLDGRLSYPAAGRAILTVRADRDVILWV
AYPLVPGPERLLVLAADAGLDYADTDSAGPAVERIAFAFALHTDFFPGADEL
ARPLPILPSMVSAGESARIYAQVLELGPVLEFSQHDVPVTPDMGPRRAERARRE
RAAALAAAGPVPDELRDEADELYAAVRERLEEFLENEVSGAIGTSLDSRTIVEVSR
VVPRTDVAGTYLREOVAGEGFPDGVPTDTMMHVRVALEHTDPRGWDVVPVGEAM
PFAHTPFQCMTTGEPVLPRIETEEMGHAIASOFEKDIPRLITGRSMLAVLPKART
VULGFMLLHHPERPVPENDRVTCASLARAGLVLDNARMYTHQENVAEITLQDSMLP
HIPARLAGCDIATRYLPGLTLLGRVGGDFDAVKLPKARTALVVGDMGHGNSAAMMG
QLRTAVQTWAGLDLPPAQLRLNDLLAQLRGDTHLATCLYAVYDPIAGELHLNAGHI
PVLVRAEDCASBLIDLPTGAPIGVGGVPPESVTVVAPGDRLVMTCTDGLVEVRGDI
GVLATLCEASAAHPAASMDACDTIIRALGAGGKKDDVALLMARLTGIEPDVAEVR
GLADPAAGRAARVREOLYDWGLPKLVAPAEILMVSELVTNAVHSHARPVALRLVRA
DTLCEVDDDDHELPAKSGAGDETGRGLRVSTLAREWGSRTGAGKTWNFELTLP
ARRH"
7711..8433
/gene="SCO6485"
/notes="synonym: SC9C7.21"
7711..8433
/gene="SCO6485"
/notes="SC9C7.21, unknown, len: 240 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC9C7.21"
/protein_id="CAA22733.1"
/db_xref="GI:4154080"
/db_xref="GOA:Q9ZBI9"
/db_xref="SPTREMBL:Q9ZBI9"
/translation="MSVTISYRDAGEFWNEASGDRGAVFWDAPFTLTAGRHAFHPEP
HLTDAALPIDLGCNGCTQTRYPAERFPHVVGADLSAAALAHASHADPGGLAAVYQLD
AADKAEQALHAEGLDANVYRGVHQCEDDROPLADALAAILGERGAEFLVRSQA
AKTVLMGLAQGAPPPAKLAPVLNKHGLTPGAVPDEAVPEYLRAAGLTVLASGELPLVT
TEYTABDQTRIELPSAWVYAGTA"
8603..9436
/gene="SCO6486"
/notes="synonym: SC9C7.22"
8603..9436
/gene="SCO6486"
/notes="SC9C7.22, possible transport associated protein,
len: 277aa; similar to SW:DPPA BACSU dppA, dipeptide
transport protein (274 aa) fasta scores; opt: 448,
z-score: 604.3, E(): 2.3e-26, (32.3% identity in 269 aa
overlap)".
/codon_start=1
/transl_table=11

```

## Alignment Scores:

Pred. No.:	1.68e-65	Length:	296500
Score:	2493.00	Matches:	480
Percent Similarity:	67.33%	Conservative:	132
Best Local Similarity:	52.81%	Mismatches:	236
Query Match:	48.55%	Indels:	61
DB:	1	Gaps:	18

US-09-917-376-1 (1-957) x SCO939128 (1-296500)

QY - 10 ThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAla 29









Db 176223 TGTGTCAGGTACACGACTTCCCGAAGCTGCGGCACCTACGTGACGAGCGGACCGACAG 176282  
Qy 227 ThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAlaPheAspLysSerSer 246  
Db 176283 AGCGGTACGCGTCCGACAAACAGGCGATCGTGGGTACCTTCGACGAGTCGACGGG 176342  
Qy 247 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnProVal 266  
Db 176343 TCGCCGGGAGCTCCACGCGAGCGGTACGTGCGGGTCCGCCAAGACAACTCCGTC 176402  
Qy 267 PheTyrSerArgAspGlyAlaThrTyrGlnAlaValProGlyAlaProThrPhePhe 286  
Db 176403 TATCGTCCACGACGCGGCGGACCTGTCGCCGTGGCCCGGACCGCCACCGGCGCAT 176462  
Qy 287 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 306  
Db 176463 CTCGCCACAGGGCGGTGCTGGACGGCGGCAACGGCTGTCGTACCTCGGCTACAGGAC 176522  
Qy 307 ThrGlyGlyProTyrAspGlySerSerGlyAspValTyrLysPheSerValThrSerGly 326  
Db 176523 AAGGGCGGACCGTACGACGGCGGCAAGGACAGCTGTGGCGGTACACGACGAAGACCGGG 176582  
Qy 327 ThrTyrThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 346  
Db 176583 ACCTGGACGAACATCAGCCCGGTGCGGAGCGCGACAC-----TACTACGGCTTC 176633  
Qy 347 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 366  
Db 176634 AGCGGCTGACCTGGACCGGAGCATCCGGGACGGTGTGCGGCTGCTGCTACAGCTCC 176693  
Qy 367 TrpTyrProAspThrIlePheArgSerThrAspGlyAlaThrTyrThrArgIle 386  
Db 176694 TGTGTCGCGACACGACGCTCTTCGCTCCACGACAGCGCGGCACCTCGACGAAGGCC 176753  
Qy 387 TrpAspTyrThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 406  
Db 176754 TGGGACTACACCTCGTATCCGAGCGCTCGAACCCTTCACCATGTGTCCTGCTCCGTC 176813  
Qy 407 ProTyrLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTyr 426  
Db 176814 CCTGTGCTACCTTGGGAGGGAACCCCGCACCGCCGAGACGACCCCGAAGCTCGGCTGG 176873  
Qy 427 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 446  
Db 176874 ATGACCGAGTCCCTGGAGATCGACCCCTTCGACTCCGCGGCGATGATGACGAACGGGC 176933  
Qy 447 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTyrAspSerGlyGlyGlnIleHisIle 466  
Db 176934 GCGACGGTCTACGCGCACGGAACCTGACCAACTGGGACAGCGGAAGCCAGTTTCAACCATC 176993  
Qy 467 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer 486  
Db 176994 AAGCCGATGGCGGGGCGCTGGAGGAGACGGCGTCAACGCTCGCCTCGCTCCCTCC 177053  
Qy 487 ---GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspVal 505  
Db 177054 GCGCGGCCCGAGCTGTTCAGCGGCTCGGTGACATCGCGGCTTCGCGCACAGGACCTC 177113  
Qy 506 ThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAsp 525  
Db 177114 ACCACGGTGGCGTGGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 177173  
Qy 526 TyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerGln 545  
Db 177174 TACCGCGAGACCGACCGGCGACGGTGGTGGCGGTGCGGCAATCTCGAC-----TCGGGT 177227  
Qy 546 ProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTyrPheGlnGlySer 565  
Db 177228 CCG-----CATGTGGGCTTCTCGACGACCAACGCGGCCCACTGGTTCGCGGGGCG 177278  
Qy 566 GluProGlyGlyValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPhe 585  
Db 177279 GACCCCTTCGGGGGTTCAGCGGGGTGGGACGCTCGCGCGCGCTCCGACGCGCAGTTCGCTTC 177338

Qy 586 ValTyrAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer 605  
Db 177339 GTGTGAGCCGCGCGGCACCGGG-----GTGCAGTACACGACCGGGTTCGGCACCTCG 177392  
Qy 606 TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnPro 625  
Db 177393 TGTGCGCGCTCCGCGGCGCTCCCGCGCGGCGGATCGTCGATCCGACCGGCGTCCACCG 177452  
Qy 626 LysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThr 645  
Db 177453 AAGACCTTCTACGGCTCAAGTCCGCGACAGTTCATCGTCACTCGAGCGGCGGCGGAC 177512  
Qy 646 PheGlnProValAlaAla---GlyLeuProSerSerGlyAlaValGlyValMetPheHis 664  
Db 177513 TTCACGCGCTCCGCGGCACCGGCTTCGCGAGCGCGACAGC-----GTGCGCTTCAAG 177566  
Qy 665 AlaValProGlyLysGluGlyAspLeuTyrLeuAlaAlaSerSer----- 679  
Db 177567 GCGTGCCTCCGCGACGAAGGCGACATCTGGCTGGCGGCGCGGAGCGACGCGCGCTAC 177626  
Qy 680 GlyLeuTyrHisSerThrAsnGlyLysSerTyrPheAlaIleThrGlyValSerSer 699  
Db 177627 GGGCTGTGGCACTCGACGAGCGCGCGCGCTTACCAAGCTCCGCCCGTCCACCG 177686  
Qy 700 AlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 719  
Db 177687 GCGGACACCATCGGCTTCGCGAAGCGCGCGCTCGTACGACGCTCTACAC 177746  
Qy 720 ValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrp 739  
Db 177747 AGCGCAAGATCGCGGTGTGCGCGCATCTTCGCGTCCGCGCACGACGCGCGAGCTGG 177806  
Qy 740 ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTyrGlyGlnAlaIleThrGlyAsp 759  
Db 177807 ACCGCGCTCAACGACGATGCCACAGTGGGTGGAGCGGCGCGCGCATCACCCTGAT 177866  
Qy 760 HisAlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyValTyrGlyAsp 779  
Db 177867 CCCAGGCTCTACGGCGCGGTGTATGTGTGCGAAGCGCGCGCGGATCGTCTACGCGAC 177926  
Qy 780 IleGlyAlaProSerGly-----SerProSerProSerValSerProSerAla 796  
Db 177927 ACCGCGGCTCTTCGCGACGCGCGGTGTACGAGCGCGCGCACGCGGCGCTGCGCGTG 177986  
Qy 797 SerProSerLeuSer---ProSerProSerProSerProSerProSerProSer 815  
Db 177987 ACATACAGGATCACGAACCAAGTGTGCGGCGGCTTCCAGCGCGGATGTACAGCTCCCAAC 178046  
Qy 815 rProSerSerSerProSerSerSer-----Pr 824  
Db 178047 ACCGGTTCGACCGCTGGGACGCTGCTGCTGGGTGCTGCTGCTGCTGCTGCTGCTGCTG 178106  
Qy 824 oSerProSer----- 827  
Db 178107 GTCAACCGAGCTGTGGAACGCTCGTACGCGCAGCGGTTTCGGGGGTGACGCGACGGAAC 178166  
Qy 828 -----ProSerPro----- 830  
Db 178167 CTGGCTTGGAAACGCGGAGGTGCGCGCGGCTCTCGGTGAGCTTCGGGTTCACGGGAAGC 178226  
Qy 831 -----SerProSerProSerProSerProSerAla-----Se 843  
Db 178227 TGTTCGGGATCCATGGCAGACCGCGCTTCAAACTCGCGGCTGCAACTGCGACGGTTC 178286  
Qy 843 rPro-----SerProSerSerProSerProSerProSerProSerProSerProSe 861  
Db 178287 GCCTGTATCGAAGCGCA-CGCTCGCTGATCCGAGGCGCGCGCGCGCGCGCGCGCG 178345  
Qy 861 rProThrProSerSerSerProValSerGlyGlyValLysValGlnTyrLysAsnAsnAs 881  
Db 178346 CCCTCCGCGCTCTCCACGCGCGTGTAGACCGTGGCGCTG-----GGCAA 178387





1156 GGGCCATATAATGGTGAATATGGAGAGATGAGATACAACTAAGACCGGAGANTGG 1215  
 329 ThrArgIleSerProValProSerThrAlaAsnAspTyrPheGlyTyrSerGly 348  
 1216 AAAAATAATAAGTCCATGGCTGCTCAAGATAC-----TATTTGGATATGGTGA 1266  
 349 LeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTyrTrp 368  
 1267 TTGGCTGTAGAGCCCGCAGAAAGTATGATGGTGGGAAACGTTGGGAA 1326  
 369 ProAspThrIlePheArgSerThrAspGlyValAlaThrTrpThrArgIleTyrAsp 388  
 1327 CCAGATACCTATATTGGGAGAGATATCGATGGTGGGAAACGTTGGGAA 1386  
 389 TrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrp 408  
 1387 TGGATGGCTATCCTTAATAGAACATTCATTAATAGATGATATTTCTGAGCTCTTGG 1446  
 409 LeuThrPheGly---ValGlnProAsnProValProSerProLysLeuGlyTyrMet 427  
 1447 CTAAACCTTTGGAAWTACTATCTACACCACTGAAAGTAAAGCCCAAACTTGGCTGGATG 1506  
 428 AspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAla 447  
 1507 GTTGGAACTTGAATAGACCCATTAATTCAGATARAATGCTTTATGGAACAGAGACT 1566  
 448 ThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAla 467  
 1567 ANATTGTATGGATGCGCATGATTTGACAAATGGGACAAAGGGCAAAATATCACCATTA 1626  
 468 ProMetValLysGlyLeuGluGluThrAlaValAlaAsnAspLeuIleSerProSerGly 487  
 1627 GTCAGGCTATAGGAATGAAGAACTTCAGTCCAGCACTGATAGCCCAACAGTTGGA 1686  
 488 AlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAla 507  
 1687 CCACATTTATCAGTGCACCTGGGAGATATAGCGGATTCAGACATGAAGACTTTGGAGAA 1746  
 508 ValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAla 527  
 1747 GCACCTTAATGGACATATGTTTCAAGCTTAATATGGGACTACTACTGATATGATTTTGT 1806  
 528 GluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSerGlnProAsn 547  
 1807 GAGTTAAATCCAAATTTATGGTTGGTGGAGGAATGTTGATAACAGTGGATCCAAAT 1866  
 548 AspArgHisAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGly---SerGlu 566  
 1867 ACTAACAGAAATGTTCTCATATGATGGTGGAAATCATGTTTCAGGGAAATACTGAA 1926  
 567 ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 586  
 1927 CCTCAAGGAACACGAAAGTGTACAGTTGCGAGCTGCTCAGATGCTGTGTGTT 1986  
 587 TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTyr 606  
 1987 TGGGCAACCA-----AAAGAGCGGAAGTATGCTATTTCTACAGTAACGGAACAAATGG 2040  
 607 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 626  
 2041 GTTGAATGCGCAAAATGTACCTTCAGAGCTATGTTATATTTCTGACAGATAAATCCAAAT 2100  
 627 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 646  
 2101 AAGTTTTATGCAATTAAGAAATGGCAAAATTTATATAGTCTGCTAAGGAAACAAATTT 2160  
 647 ---GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAla 665  
 2161 ATCGAAGGCTCTGCTGCTGTTTACCAATATCAGGA-----AATTCAAACT 2208  
 666 ValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThr 685

2209 GTTCCAGGTATAGAGGTGATATCTGGCTAGTTGGTATAATGGATCGGACCTCAACT 2268  
 686 AsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPhe 705  
 2269 GATGGTGGATATTCATTTGTAAAAATTTCAAGTGTGAGGATCGACGAGTATAGATTT 2328  
 706 GlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGly 725  
 2329 GGTAACCTCTCTGAAGAGAAACATATCTCTCAATTTATCTTATGCAAAAATTAATGA 2388  
 726 ValThrGlyAlaTyrArgSerAspAspCysGlyThrTrpValLeuIleAsnAspAsp 745  
 2389 GTAAGAGTATATTTAGATCTGATGATGATAAAAGTGGATAGAAATAAATACGAC 2448  
 746 GlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg 765  
 2449 AAACATCAGTTTGGTCTGCGAATGCAGACATAACTGGAGATCCACGGGTGTACGTCGA 2508  
 766 ValTyrIleGlyThrAsnGlyValTyrGlyValTyrGlyAspIleGlyGlyAlaProSer 785  
 2509 GTATTTTGTGCAACAAACGCGCTGGGTATAAATGGGGTGAATTTGCT-----TAT 2559  
 786 GlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerProSer 805  
 2560 TCAATATTTGCCATCAGCACTCCACACTCAACATCAACCAAGTTTCACCCGACCGACT 2619  
 806 ProSerSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 825  
 2620 CCAACGCAACTATTATCATCCACACTGCACCAACATCAACCAACCAACCAACCCGACC 2679  
 826 ProSerProSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 845  
 2680 CCAACCAACAGTGACGCAACGCGCACGCGCACGCGCTTACAGTGACGGTTACT 2739  
 846 ProSerSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 865  
 2740 ATGACTCCGACCCGACCA-----ACACCAACGCGCGACCAACCA 2781  
 866 SerSerProValSerGly---GlyValLysValGlnTyrLysAsnAsnAspSerAlaPro 884  
 2782 GGGACACTGTGACGGGAGTGGTTTGAAGTACTGTACAGAACATCAGACGAGTGG 2841  
 885 GlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySerSerSerValAsp 904  
 2842 AGTCAGGTTCTATAAGCGCGTGGTTAAGATAGTAGTAATGGAGGCGACGAGTGTGAT 2901  
 905 LeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuVal 924  
 2902 CTTAGCAGGTTAAGATAAGATCTGTTACAGTGGATGGTGTGACAGCAAGCCAGAGTGG 2961  
 925 TyrAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerVal 944  
 2962 GTA---TGTGACTGGGACAGATAGTGTCAAGCAATGTGACATTCATTTTGTGAAGCTT 3018  
 945 AsnProAlaThrProThrAlaAspThrTyrLeuGln 956  
 3019 AGCAGCGGAGTGGAGCGGATTTATTACTTGGAG 3054

RESULT 5  
 AX565635  
 LOCUS  
 DEFINITION Sequence 1 from Patent WO02077242.  
 ACCESSION AX565635  
 VERSION AX565635.1 GI:26000969  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 1  
 Duffner, F. and Sjoeholm, C.  
 Family 74 xyloglucanases  
 Patent: WO 02077242-A 1 03-OCT-2002;  
 Novozymes A/S (DK)

FEATURES		Location/Qualifiers		Db	
source		1..2823		Qy	
		/organism="unidentified"		Db	
		/mol_type="unassigned DNA"		Qy	
		/db_xref="taxon:32644"		Db	
		/note="Jonesia sp., DSM 14140"		Qy	
CDS		1..2823		Qy	
		/note="unnamed protein product"		Db	
		/codon_start=1		Qy	
		/protein_id="CAD58415.1"		Db	
		/db_xref="GI:26000970"		Qy	
		/db_xref="REMYREBL:CAD58415"		Db	
		/translation="MLRIVVATALLATATTAGGAATASTPMVTPAPSPSPSGY SWSNIVGGVFPVTFVFKQDPLGVYARTDGRVYRLANDSTGRWTLPLTDHIGWDDMS HSGIILSLATPDVDTNRVLAAGTYSMWDPPQNGAIIKRSADKSTWTKMLPFRVGNM PKRGMRVADPNNNKLVYFGAESGNGLWKSVDYGTWCKVTSFPAAGNVADASGA YTGQNGVWVDFPTSAKAGTQTLIYGVADKQNNVRSITDGGATWQVPCQPTGF LAQGVLDKQQLYIATSDTGTGPDGSGDGVWRDLSSQWTRISPISTSNAPF YSLGLADRNKPDPTIMVQVSWMPDMVYVRSITDRKTSPIWELNGSQRTKQNYHDY SGAPWLDFTNKAKEPEANPKLGMWTSFIDPHNSDRFFYTGAGIYGGTNLTNWDK KKVDITVKAQIEBTAQDLAAPPGNIGLYSALADIGFTKDIISOVPKNYKKNPH DVTSTIDFAESKPAVTRAKSISGRTTSMVGVSTDAGETWKPATPSGVKPGSLTV SANASIVNAPGAAPRRSTNSGSSWTSVGLPHNAQVADRVNANTLIFGVDCKFYH STNGGASFTASAPFTSGNVRFRVAPQGHVLAGVSGSTGYGMRSTDDGKNWT KVGVGQDGAAGFGKATSSGSPVIFTSAKIDGVRIFRSDDEGKTWKRINDNHNQWY WTGASTIGDPDVEGRVITGNRGRIIVGDSSTPPPPQDSGASGNTGATPTPETDS DINTGSGTDTGTSRPTGSGGCTAPDSGNTSPGNTSSCKVYSTTDMGSGFTGAATI TNTSPSPINGWTLAFTYPSGQTISSVWSATQTLISGRNVVLKNSGNWPTTIPAGGSVEFG FNAGYTSKNTPTTSPFSLNGVTCALAD"		Qy	
sig_peptide		1..69		Qy	
mat_peptide		70..2820		Qy	
		/product="unnamed"		Qy	
ORIGIN				Qy	
Alignment Scores:				Qy	
Pred. No.:		8..836-58		Qy	
Score:		2174.00		Qy	
Percent Similarity:		59.48%		Qy	
Best Local Similarity:		44.79%		Qy	
Query Match:		42.34%		Qy	
DB:		6		Qy	
		US-09-917-376-1 (1-957) x AX565635 (1-2823)		Qy	
		10 ThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAla 29		Qy	
		13 ACCGTGGTGGCCACAGCACTAACTGCGCTGCTCGCCACGACGTCGCGGTGCATCGCGCC 72		Db	
		30 AlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThrThr 49		Qy	
		73 ACATCA-----ACCCCATGCCAGTAACCTGACCGCCCTCTTCTCATCATCACT 123		Db	
		50 GlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleVal 69		Qy	
		124 TCCGGCTATAGCTGGTTCGAACGTTGAGATTGTGGCGGGGTTTGTGCGCTGGAATTGTC 183		Db	
		70 PheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArg 89		Qy	
		184 TTTAACCAAGAACACGAGCTTGGTGTATGCCCGCACACAGATATGCTGGAGCGTACCGG 243		Db	
		90 TrpAspAlaAlaAsnGlyArgTTrpIleProLeuLeuAspTrpValGlyTTrpAsnAsnTrp 109		Qy	
		244 CTCATGACTCCACCGGGGTTGGATCCCTCTTACCGACCATATATGTTGGAGTACTGG 303		Db	
		110 GlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAla 129		Qy	
		304 AGCCACTCCGGAATTCTGCTTGGCCACTGACCTCTTGACACCAACCGCGGTACTCT 363		Db	
		130 AlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSer 149		Qy	
		364 GCAGAGGACGCTACTCCATGACTGGGATCCACAAACGCTGCTATTTTGGCTTCGCG 423		Db	
		150 AspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLeuGlyGlyAsnMetPro 169		Qy	

424	GACAAAGCGGAGACCTGGGAAAAGACGATGCTGCCGTTCGGTGGAGAACATGCGCG	483
170	GlyArgGlyMetGlyGlyuArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPhe	189
484	GTCGCGGGATGGGAGAACGCTCGCCGTTGACCCACCAATAATAAAGCTCTGTACTTT	543
190	GlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGln	209
544	GGTGGGAAATCGGGTAATGGACTATGGAAGTCAACTGACTACGGGAAAACCTGGGGCAA	603
210	MetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGly---	228
604	GTGACGAGCTTCCCAATGCTGGTAACACTACGTCGCC-----GATGCTTCTGTGCGG	654
229	TyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeu	248
655	TACACCGGGGAAAACCAAGGTGTGTGTGGGTAACTTTTGACCCCATCAGCGAAAGCT	714
249	GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrp	268
715	GGTCAGACAAACACAAACGATCTACGTAGGTAGCAGACAGCAGACAAACGTTTATCGA	774
269	SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIlePro	288
775	TCTACCATGTCGGGGCAACATGTCGACGCGTCCCTGGACAAACCCACCGGATTCCTGGCG	834
289	HisLysGlyValPheAspProValAsnHisValIleLeuTyrIleAlaThrSerAsnThrGly	308
835	CAAAAGGGAGTGTGACCATAAAGGTCAGCAACTCTACATCGCTACCTCAGATACTGGT	894
309	GlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrp	328
895	GGCCCGTATGACGGGTGGAAGGGGATGTGTGGCGCTAGATATTTCCAGTGGCCAGTGG	954
329	ThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGly	348
955	ACTCGAATCAGCCCAATTCCTCTACC--TCATCGAATTCAGCGTTCGGGTACTCCGCT	1011
349	LeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrp	368
1012	TTAGCCATTGACCGCAAAAACCTTGACACCATCATGTTGTTTCCCAAGTCTCTGGTGG	1071
369	ProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAsp	388
1072	CCTGACATGATGATATATGCTCCACCGCGGGGAAAACGTGGTCCGCTATCTGGGAA	1131
389	TrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrp	408
1132	CTGAACCGTTCCACAGCCGGTACAAAGCAGTACACACGATTACTCAGGGGCTCCCTGG	1191
409	LeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrpMetAsp	428
1192	CTTGACTTTGGCAATACAGCCAAAGAGCCTGAAGCTAACCCCAAACTCGGGTGGATGCC	1251
429	GluAlaMetAlaIleAspProPheAsnSerSerArgMetLeuTyrGlyThrGlyAlaThr	448
1252	CAGTCTTTCGAGATTGACCCGCATAACTCCACCGCTTCTTCTATGGCACCGCGCGGC	1311
449	LeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaPro	468
1312	ATCTATGGCGGAACCACTCACAACTGGGATAGGGGCAAAAAGATTGATATCACTGTC	1371
469	MetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProSerGlyAla 488	
1372	AAGGCCCAAGGATCGAAGAAACCGACGCCCAAGACCTTGGCGCTCCCTCGAAACATT	1431
489	ProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal	508
1432	GGCCTATATCTACGCGCTGGCAGACATCGGTGGCTTCAACCCCAAAAGATATCTCTCAGGGT	1491
509	ProSerThr---IlePheThrSerProValPheThrThrGlyThrSerValAspTyrAla	527

Db 1492 CCTAACAGTACTACTACAGAACCCACACCATGACACCGTGACCTCCCTGACTTCGCC 1551  
 Qy 528 GluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSerGlnProAsn 547  
 Db 1552 GAAAGCAACCGCAACTGTCGTGCTGCGAGGCAAA-----TCTATTTCTGGGGAACC 1605  
 Qy 548 AspArgHisValAlaPheSerThrAspGlyGlyValAsnTrpPheGlnGlySerGluPro 567  
 Db 1606 ACCTCATGGGTTTCAACTGATGCTGGCGAAACGCGGAAACCTGGCGCAACGCCA 1665  
 Qy 568 GlyGlyValThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTyr 587  
 Db 1666 TCGGGGTTAAAGCCCGCCGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1725  
 Qy 588 AlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAla 607  
 Db 1726 GCCCGG-----GAAGTGCGGCTCTCGAGGTTCCACCAACAGTGCTCTCTGCTGCTCA 1779  
 Qy 608 AlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThr 627  
 Db 1780 ACGGTGAGCGGGCTCTCTCACAATGCGCAGTGCATCAGACCGAGTCAACGCTAATACT 1839  
 Qy 628 PheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGln 647  
 Db 1840 CTTTACGGTTCCTGACGGAAGTTCTATCATCTCCACAATGCTGGGGCTTCTCTCAGC 1899  
 Qy 648 ProValAla---AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 666  
 Db 1900 GCGTCAGTTTCACGGGTTTCCACACAGTGA-----AAGTTCGTTCCGTCGCGTG 1953  
 Qy 667 ProGlyLysGluGlyAspLeuTrpLeuAla-----AlaSerSerGlyLeu 681  
 Db 1954 CCGGGCGCTCAAGTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2013  
 Qy 682 TyrHisSerThrAsnGlySerSerTrpSerAlaIleThrGlyValSerSerAlaVal 701  
 Db 2014 TGGCGGTCCACTGATCGCGGAGAACTGACCAAGGTATCTGGAGTTCAGGAAGCGCAT 2073  
 Qy 702 AsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGly 721  
 Db 2074 GCGGTGGTTTCGTTAAGCAACAGTAGTAGGATACCGAGTATTTTACATTCGCC 2133  
 Qy 722 ThrIleGlyValThrGlyAlaTyrArgSerAspCysGlyThrTrpValLeu 741  
 Db 2134 AAGATTGATGGAGTTCGAGGAATTTCCGCTCTGACGATGAAGCAAGACCTGGAACGC 2193  
 Qy 742 IleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAla 761  
 Db 2194 ATCAACGACACACCAACACAGTGTATTTGAGTGGCGCATCAATCACCGGTGACCCAGAT 2253  
 Qy 762 AsnLeuArgArgValTyrIleGlyThrAsnGlyValGlyValTyrGlyAspIleGly 781  
 Db 2254 GTCAGAGGCGGCTCTACATGTGTATCCACGTCGCGGCAATTCGTTGGTGAT----- 2307  
 Qy 782 GlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSer 801  
 Db 2307 ----- 2307  
 Qy 802 ProSerProSerProSerSerProSerProSerProSerProSerProSerProSer 821  
 Db 2308 -----TCCTCCACTCCCGCCCGCCCGCTGCAAGATAGCGGCGCATCG 2352  
 Qy 822 SerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840  
 Db 2353 GGAATACAGGGGCCACACCGCTGAAACCCCTGACAGCGATCAACACATGGCGGG 2412  
 Qy 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSer 860  
 Db 2413 TCCACTGATACGGCGGCAACTTCAGACCGGGACAGGAGTGGATCGTGGNACCGCACT 2472  
 Qy 861 SerProThrProSerSerProValSerGlyGlyValLysValGlnTyrLysAsnAsn 880  
 Db 2473 GACTCGGGGAACCTCTCTCTGGAATACCACTGTTTCATGCAAGGTTCTGTTATTCACGACA 2532

Qy 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900  
 Db 2533 GACTGGGGCAGCGC-----TTCACTGGGGCAGCGACCATCACTAACACGCTCCCC 2583  
 Qy 901 SerSerValAspLeuSerThrValThrValArgTyr----- 912  
 Db 2584 AGCCCAATTACGGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2643  
 Qy 913 ---TrpPheThrArgAsp-----GlyGlySerSerThrLeuValTyrAsnCysAspTrp 929  
 Db 2644 GTGGGTGACGCAACTCAGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2703  
 Qy 930 AlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 949  
 Db 2704 -----AATCCAACGATACCT 2718  
 RESULT 6  
 AE007608 12732 bp DNA linear BCT 27-JUL-2001  
 LOCUS Clostridium acetobutylicum ATCC824 section 96 of 356 of the  
 DEFINITION complete genome.  
 ACCESSION AE007608 AE001437  
 VERSION AE007608.1 GI:15023819  
 KEYWORDS  
 SOURCE Clostridium acetobutylicum  
 ORGANISM Clostridium acetobutylicum  
 Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 Clostridium.  
 REFERENCE 1 (bases 1 to 12732)  
 Nolling, J., Breton, G., Omelchenko, M.V., Markarova, K.S., Zeng, Q.,  
 Gibson, R., Lee, H.M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y.I.,  
 Tatusov, R.L., Sabatthe, F., Doucette-Stamm, L., Soucaille, P.,  
 Daly, M.J., Bennett, G.N., Koonin, E.V. and Smith, D.R.  
 Genome sequence and comparative analysis of the solvent-producing  
 bacterium Clostridium acetobutylicum  
 J. Bacteriol. 183 (16), 4823-4838 (2001)  
 21359325  
 11466286  
 REFERENCE 2 (bases 1 to 12732)  
 Childress, D., Zeng, Q. and Smith, D.R.  
 Direct Submission  
 Submitted (24-JUL-2001) GTC Sequencing Center Production,  
 Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100  
 Beaver Street, Waltham, MA 02453-8443, USA  
 FEATURES  
 Location/Qualifiers  
 1..12732  
 /organism="Clostridium acetobutylicum"  
 /mol\_type="genomic DNA"  
 /strain="ATCC 824"  
 /db\_xref="ATCC:824"  
 /db\_xref="taxon:1488"  
 73..2592  
 /gene="CAC0919"  
 73..2592  
 /gene="CAC0919"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Probably secreted sialidase; several ASP-boxes  
 and dockerin domain"  
 /protein\_id="AAK78895.1"  
 /db\_xref="GI:15023820"  
 /translation="MNKRIIVSMVAGLSITFTTFVTHISAAANKAAASVQSCYKWDN  
 AKIGAGGVPAVFNKTEKDIYARTDMGGAYRWKANNKWIPIIDGFSDFMTLGCES  
 IATPIDTRNVIIAAGLTNDWQDENATILSSQDKNTGNTWKRLQLFPKVGNNPGRNMG  
 ERQIDINDNKILYLGARSNGLSKSDYQSDVDFPDFTGVDVQDFQNEYTADKV  
 GVWETDPSTGTSGSPQTYVGAADTKNNIYVINDGKTWSAVKQPGKGLPHHG  
 ILASDGLTYSISNTSGPDGSGQVKNYKTGWTNITPPAVGDTKSGFGSIVDA  
 ONPNVVATLNRWNPDEELIYRSTDAKTWKPIWNGYFNRTLGVNLIDYSAQPLDW  
 KGTGVTPDPLVKGMMGDLIEIPFNSDRPFYGTGALYGTDDLTDNWDKGNVLSV  
 KANGIEECANVDVVVPTGALLSAGDDCGFYHDDITKVPKSMITTNFSAITSLDY  
 AESVNFVVRVGVNDTSKNQDKCGISYDGKKNWFSAGSNTSGYTKAGTVAAGDAK  
 TIVWSPBEGANAAYSTDNKNTWPCSLGLPQGAKVRSRDRVNPKKFYGLNGKFYSTDA

gene  
CDS  
GATFTOSSQTLTKGKGIPTKTVLHGGDIWAGKGLWHSTDSGATFTKVSQVDAS  
DTVLGKSKIDDGPALYMDATIDGTAGIFRSDDGATWVRINDDAHQYSGPYCITAG  
DPNKYGRVFGTNGRIGVIGDTSQPTPTSPVTLGVDNDGDIINGRDVNVLTQYTAG  
KQVTDIKAADVNGDGTNGRDLMTLRQLYISGRITSPSTK"  
complement (2634. .2978)  
/gene="CAC0920"  
complement (2634. .2978)  
/gene="CAC0920"  
/codon\_start=1  
/transl\_table=11  
/product="Protein related to MIFH/DOPD protein family,  
function in bacteria is unknown"  
/protein\_id="AAK78896.1"  
/db\_xref="GI:15023821"  
/translation="MPFINSTVTVKLDKEKSTLKVGLKGIIELLPKSDDWLVGFK  
DNPVLFYKGEKQKAAFIKVGISADKASKNKLVAISSIMENELSPKDCIVITVD  
EVSNNWAWNGSLP"  
3104. .4252  
/gene="CAC0921"  
3104. .4252  
/gene="CAC0921"  
/codon\_start=1  
/transl\_table=11  
/product="Metal-dependent amidohydrolase"  
/protein\_id="AAK78897.1"  
/db\_xref="GI:15023822"  
/translation="MITIEKQBEFLKLIGYRELKHHPSLMKYETTKRIKRWLVE  
NDIKVLDPTELVGVAVIEGEHEGKTVARADIDALPIEKTLPSSVNVKGMHACG  
HDPHTAALIGTALILSKRAEIVGRVRIIPQGEETGKSGOYIEGVLRVDVCIFGM  
HNKPDPVGTIIRSGELMASVDRFEIKVIGSGHAGIPNKCDIPVVASQIVASQMT  
VVSRSIPSTESVITRFSNGTNNVSDSMEGTFTNNIRIEEIRNLMERTAK  
GIARALGKISGFVLENDVRFKDCIEAEDNGYVVEAKRNILGGEDPSFYQ  
TVPSFPVWVGVDGTEBWHKANYNLKEDALMVAANYPSDLALAKVLSK"  
4594. .4863  
/gene="CAC0922"  
4594. .4863  
/gene="CAC0922"  
/codon\_start=1  
/transl\_table=11  
/product="Uncharacterized small conserved protein, homolog  
of yfjA/yukE B.subtilis"  
/protein\_id="AAK78898.1"  
/db\_xref="GI:15023823"  
/translation="MFDKAALECEDIKSIKSTTOLDSWYGESCTAPEKTVILDR  
KLNTYTEILQEYSKALYDISETYQNTGDIKAKAMKAVNEASEQ"  
5056. .6348  
/gene="CAC0923"  
5056. .6348  
/gene="CAC0923"  
/codon\_start=1  
/transl\_table=11  
/product="Predicted membrane protein; CF\_1 family"  
/protein\_id="AAK78899.1"  
/db\_xref="GI:15023824"  
/translation="MOSLQVDRNGLLKINEAKOGLKSAVDKPSLKNKRSRIELKN  
LRELKAVNOLGEIRKINSNDIMYKRTDVKKCAKSLKTGYTYRKKVGLNFGH  
EVERPVIGIGVPIDAVESWDRNKFVLKTTITITETGAIGVCIVITPEAGITWKA  
WDLAAGVNSLDNIGTASAIYQKRVKSESDSSGVDLFGSIFGGAGWSAKTLG  
VKNPSYSGSGTXYDRTSATVTLNVGHSVTSLGDKVEKLPALKKAEASANKTLGA  
ANNAEKIHMDAMATKINGINNAKEVVDLKKKNVKTNFEPYKQFYKDKNLL  
KDKGNKLPOLOQDLIIQDQLETFEKNITLNAKNYKNSIRDLNKNYRDIKASLYNIPA  
GGTLNTEVDQGSFHFNYFGEIKPIFEIKVVLGGER"  
6345. .6923  
/gene="CAC0924"  
6345. .6923  
/gene="CAC0924"  
/codon\_start=1  
/transl\_table=11  
/product="Predicted membrane protein"  
/protein\_id="AAK78900.1"  
/db\_xref="GI:15023825"  
/translation="MIRFOKRWAIPLFTISILIIQGDIEKIIISFGIEFDLILCNLPFI  
GLGIIGFIRYRWKQIKKYTELLKEDRLDKIEIYDIERYRGVVYNTAFYIDGVK

gene  
CDS  
KLRFPPQYKCKDNNDYVYLLPIYKNGEFKDFIAYADKNDYKKSIESITNEITECEEK  
AEKKRKRRLNEYVNSRMELAKKYEERAKNM"  
7003. .8043  
/gene="CAC0925"  
7003. .8043  
/gene="CAC0925"  
/codon\_start=1  
/transl\_table=11  
/product="TPR-repeat-containing protein"  
/protein\_id="AAK78901.1"  
/db\_xref="GI:15023826"  
/translation="MENENVLRQAQLNNRGLIFERNAGDEALKYFNKAIKEDNPFKE  
AYLNAALCLAMNVDDAVCYDKLIIKYPNEKAYFGKGNVFFTKNIIKAIIEUYN  
RSYILGMSGVYNNLALCMESIGELDDAIKWLNRALINDQNIITLNKKAVALIKIK  
KFTYALSCYRNLKIEVDNERVYHFKAMLLGEMKIEKAFETISIGEALLGQMTILCY  
DKAVLPKGEFERALECTERALLPDEGNVLLILKKGLLTVLKIDKASTVDKILE  
FHNMEGVFAKANILCMLGHEHDAEBEFFEILIERLSGEESVLINSYVYRALNLKWA  
KTRKQKRLIRKQ"  
7998. .8357  
/gene="CAC0926"  
7998. .8357  
/gene="CAC0926"  
/codon\_start=1  
/transl\_table=11  
/product="TPR-repeat-containing protein"  
/protein\_id="AAK78902.1"  
/db\_xref="GI:15023827"  
/translation="WGNKEEAKAYKEAIKKYNELIKYPVDVKLNFFKANCRLDIED  
YKABELEYIIDLNDALALEVYLMARNRISLNKYEAAKEDLNTVTKLNPAYRKVIEL  
DEQLKEFLSGKSECYR"  
8379. .8675  
/gene="CAC0927"  
8379. .8675  
/gene="CAC0927"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein, CF-2 family"  
/protein\_id="AAK78903.1"  
/db\_xref="GI:15023828"  
/translation="MNSDEIRIDEKACDDVIKDKAVGEKVKETYNLHLSNRHIEDG  
IKGSVDSITTEYNNKIVKMKEMQDNMGNDVAAVNNIVSFSDAEDKMIKTQYR"  
8713. .9012  
/gene="CAC0928"  
8713. .9012  
/gene="CAC0928"  
/codon\_start=1  
/transl\_table=11  
/product="Uncharacterized small conserved protein, homolog  
of yfjA/yukE B.subtilis"  
/protein\_id="AAK78904.1"  
/db\_xref="GI:15023829"  
/translation="WVDRCDIEIDASIFETIKVYSDSKVRLHDLCDLENELKQMED  
TWEQDAKKEPDSFPFGFYSAKKOCDMLEELIKLVKTSFESLDTMKNLDDK"  
9712. .10335  
/gene="CAC0929"  
9712. .10335  
/gene="CAC0929"  
/codon\_start=1  
/transl\_table=11  
/product="SAM-dependent methyltransferase"  
/protein\_id="AAK78905.1"  
/db\_xref="GI:15023830"  
/translation="WGSNQIPDNIAASNVIRSEYFELRYKVLKSVISKDKITADF  
GCGTGISLALADANNVGLDNNMLRELKSAHDKLNNIYLLKSLTSVTLFDE  
SIDAFINMALHHIKDAEKAIKENYRVLKDGTVVSDVSEHNGWAREMEPDELGF  
SNEQIDFWLKNVGFKNIDIENTDLKCKGTSRSGEHTSTSIPIAKAAK"  
10353. .11498  
/gene="CAC0930"  
/note="metB"  
10353. .11498  
/gene="CAC0930"  
/codon\_start=1  
/transl\_table=11

326	GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaIleAspTyrPheGly	345
1000	GGAGAATGGCAACAACATAACTCCACCTCGCAGTAGTAGTGACACTAAATCT-----GGT	1050
346	TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle	365
1051	TTTGGGAGGAATATCATGTAGTATGACAAATCCAAATAATGTTGTAGTTGCACATTAAAT	1110
366	SerTrpTrpProAspThrIleIlePheAArgSerThrAspGlyGlyAlaThrTrpThrArg	385
1111	AGATGGTGGCCGATGAAGAAATATATACAGAAGTACACAGATCAGTAAGACCTCGAAACCT	1170
386	IleTrpAspTrpThrSerTyrProAsnArgSerIleuArgTyrValIleuAspIleSerAla	405
1171	ATATGGGATTTGGAATGGATATCCAAACCGTACTTTGGGATATAACTTAGATTTACTCAGCA	1230
406	GluProTrpLeuThrPheGly---ValGlnProAsnProValProSerProLysLeu	424
1231	CAACCTGGCTTGGATTGGGTGAAGCCGGGTAAACACCACGACATCCAATTAGTTAAATTA	1290
425	GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGly	444
1291	GGCTGGATGATGGCAGACTTAGAAAATTGATCCCAATTAACTCGGATAGAATGTTCTACGGT	1350
445	ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrIlystTrpAspSerGlyGlyGlnIle	464
1351	ACAGGTCAACTCTTTATGGAACTGTGATTTGACTAAATTTGGGATAAAGTAAAGATGTA	1410
465	HisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerPro	484
1411	GACATCTCTGTAAGGCAATGGAATTTGAAGATGTGCAGTAAATGATGTTGTAGTACCA	1470
485	ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAsp	504
1471	ACAAAAGCGCACAACTTTTAAAGCGCTGTTGGAGATGACTCGCGATTTTATCATCATGATGAC	1530
505	ValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerVal	524
1531	ATAACTAAAGTTCCAAGTAAATGATGACAAACCTTAACTTTGTGCTACAAACAAGCAT	1590
525	AspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSer	544
1591	GATATCGCAAGAGTGTCTCACTTTGTTGTTTCGTGTTGGAAATGTTGATAGTATAA	1650
545	GlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln---	563
1651	AATCAACAAGATAAAGATTGTGGAATTTCTACGATGCTGGTAAAGAACTCGTTCTCAGCT	1710
564	GlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySer	583
1711	GGAAAGCAATATATCAGAGGTTTATAAGCAGGAACCTGTACAGCTGGTGTGATGCTTAAA	1770
584	ArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGly	603
1771	ACTATAGTAGTGTCTCCA---GAGGAAGGTGCTTAATCAGCATATAGCACAGATAATGGT	1827
604	AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal	623
1828	AATAAATGGCAACCGCTGCTCGGTCTTCCACAGGAGCAAAAGGTAGATCAGACCCGTGTA	1887
624	AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly	643
1888	AATCCAAGAAATTCATAGTTTCTTAAATGGTAAATTCATATATAGTACTGATGACAGT	1947
644	ValThrPhe---GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMet	662
1948	GCAACATTTACTCAAGATTCAACAAACAGGTCTCTCTCAAAAAGGCAAAA---GGAATC---	2001
663	PheHisAlaValProGlyLysGlyGlyAspLeuTrpLeuAla---AlaSerSerGlyLeu	681
2002	TTTAAAGACTGTTATAGTCATGAAGTGATGATTTGGATTTCTGCGAAGAAAGACGGAATTA	2061



Qy 682 TyrHisSerThrAsnGlySerSerSerSerSerAlaIleThrGlyValSerSerAlaVal 701  
 Db 2062 TGGCAATCAACTGATCTGCTGCAACATTCACAAAGTATCTGCTGTGATCTCTCAT 2121  
 Qy 702 AsnValGlyPheGlySerSerAlaProGlySerSerSerProAlaValPheValValGly 721  
 Db 2122 ACATAGGTCTTCGAAATCCAAACAGATGATGATGATGATGATGATGATGATGATGAT 2181  
 Qy 722 ThrIleGlyValThrGlyAlaTyArgSerAspCysGlyThrThrThrThrValLeu 741  
 Db 2182 ACAATTGATGGAACACGGGATATTCCTGCTCAGATGATGATGATGATGATGATGATGAT 2241  
 Qy 742 IleAsnAspAspGlnHisGlnTyArgSerThrGlyGlnAlaIleThrGlyAspHisAla 761  
 Db 2242 ATAATGATGATCTCACCAATATGTTCCAGACTTACTGATGATGATGATGATGATGATGAT 2301  
 Qy 762 AsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGly 781  
 Db 2302 AAATATGACGAGCTTTTGTGGAACATATGGAAGAGTATGTTTATGATGATGATGATGAT 2361  
 Qy 782 GlyAlaProSerGlySerProSerProSerValSer 793  
 Db 2362 GGTCT---CAACTACACCTACCAAGGTACT 2394

## RESULT 7

AJ585344 2950 bp DNA linear BCT 11-OCT-2003  
 LOCUS Clostridium thermocellum xghA gene for beta-1,4-xyloglucan  
 DEFINITION hydrolyase, strain F7.  
 ACCESSION AJ585344  
 VERSION AJ585344.1 GI:37651952  
 KEYWORDS beta-1,4-xyloglucan hydrolase; xghA gene.  
 SOURCE Clostridium thermocellum  
 ORGANISM Clostridium thermocellum  
 Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 Clostridium.

## REFERENCE

1 Zverlov V.V.  
 New cellulosome components of *C. thermocellum*  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2950)  
 Zverlov V.V.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-OCT-2003) Zverlov V.V., Fachgruppe f. Mikrobielle  
 Mikrobiologie, TU Muenchen, Am Hochanger 4, Freising, 85350,  
 GERMANY

## FEATURES

source  
 1..2950  
 /organism="Clostridium thermocellum"  
 /mol\_type="genomic DNA"  
 /strain="F7"  
 /db\_xref="taxon:1515"  
 346..2887  
 /gene="xghA"  
 346..350  
 /gene="xghA"  
 359..2887  
 /gene="xghA"  
 /note="belongs to GH74 and contains a cellulosomal  
 dockerin module"  
 /codon\_start=1  
 /evidence=experimental  
 /transl\_table=11  
 /product="beta-1,4-xyloglucan hydrolase"  
 /protein\_id="CAE51306.1"  
 /db\_xref="GI:37651953"  
 /translation="WVKFTSKIAVFAAVVAATAI FGAISSQAVTSVPYKNDNV  
 IGGGGFMPIGVNTEKDLI VARADIGAYRWDPTETWI PLLDHFQWDEYSYGYVE  
 STATPDVDRIVYIVAGMTNDLNMGAIALRSTDRGETWEKTLILFRKNGMPPGRSM  
 GERLAIDPNDRILYLGTRCGNLWRSTDYGTWSKVESFPNPGTYIDPNFDTYKDI  
 IGWVWVFDKSSSTRGNPKTIYVGVDKNESIYRSTDGSGVTWKA VPGQPKGLPHHG  
 VLASNGMLYITGDTGCPDVGNGKGVKWFNTGTGEWIDITPIPIYSSSDNRPCFAGLA

## gene

## RBS

## CDS

## ORIGIN

Alignment Scores:  
 Pred. No.: 9.86e-53 Length: 2950  
 Score: 2007.50 Matches: 402  
 Percent Similarity: 64.11% Conservative: 107  
 Best Local Similarity: 50.63% Mismatches: 250  
 Query Match: 39.09% Indels: 35  
 DB: 1 Gaps: 17

US-09-917-376-1 (1-957) x AJ585344 (1-2950)

Qy 11 MetArgSerArgArgLeuVal-----SerLeuLeuAlaAlaThrAlaSer 25  
 Db 341 TTAAGGGAGGAGCGCTTAATGGTTAAAGAGTTTACAGTAAATAAAGCTGCTGTTTT 400  
 Qy 26 PheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHis 45  
 Db 401 CGGCTGTGTTGCTGACCGCATATTTGGCCCGCGGATTTCC-----AGCCAG 451  
 Qy 46 AlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPhe 64  
 Db 452 GCTGTAAACGAGCGCTTACAAATGGGACACGCTGTAATCGGGGAGCGGAGGATTT 511  
 Qy 65 ValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIle 84  
 Db 512 ATCGCGGTATAGTTTTTAATGAAACGGAAGGATTTGATTATGTCACGTCGCCGATATC 571  
 Qy 85 GlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTyrPileProLeuLeuAspTrpVal 104  
 Db 572 GGAGGAGGTACCGGTGGGATCTTCGACCGAGCATGGATTCGTTGCTGGACCATTTTC 631  
 Qy 105 GlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThr 124  
 Db 632 CAATGGATGATGATACAGTTATTCGAGTGGAAAGTATTCACACCGACCTGTGGATCCG 691  
 Qy 125 AsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAla 144  
 Db 692 AACCGTGTTCATAGTTGACAGTATGTATACCAACGATTTGGCTTCTTAATATGGGAGCA 751  
 Qy 145 IleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeu 164  
 Db 752 ATTTCTTCGCTCAACGGACAGGGAGAAACATGGGAAAAAACCATATCTGCTTCAAGATG 811  
 Qy 165 GlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAsp 184  
 Db 812 GCGGAAACATCGCGGAGAGATCCATGGGAGAACGCTCTTGGATCGACCCGATGACAC 871  
 Qy 185 AsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGly 204  
 Db 872 AGGATTTCTTTATCTTGGAAACACGATCGGAAACGACCTTTGGAGAAAGTACCACACTAC 931  
 Qy 205 AlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThr 224  
 Db 932 GTAACATGGTCCAAAGTTGAAAGTTTCCCAATCCCGGAACCTTACATTTATGACCCG 988  
 Qy 225 AspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSer 244  
 Db 989 ---AATTTTGATTTACCAAGACATTTATCGAGTAGTCTGGGTTGTTTTCGACAGAGC 1045  
 Qy 245 SerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsn 264  
 Db 1046 AGCAGTACACCGGGCAACCCCTACCAAGACTATATATGTTGGTGTGGCTGATAAAACGAA 1105

VDRQNPDIIMVTSNNANWPPDEYIERSTDDGGATWKNINWEMTPEIRILHVEIDISAAPW  
 LDWTEKQLPEINPKLGMIGDIEIDPNSDRMMYVTGATTYGCNLTLDWRGGRVKI  
 EVKATGIEECALVLDLSPPEGAPLVSAVGDVLFVHDDLKVGKQKHVPSYSGTGID  
 YAEIYLPNFMALVAKADLYDVKKISFYDGRNWFQPPNEAPNSVGGGSVAVAADKSV  
 IWTPEASPAVTTTDDGNSWKVCTNLGMAVVASDRVNGKKFYAFYNGKFYISTDGLT  
 FDTDKAPQLPKSVAKIKAVPQKEGHVWLAAREGLWRSTDGGYTEPKLSNVDTAHVVG  
 FGKAAPOGDMYAIYITGKIDNLVGFRRSDDGKTVWRINDDHGGYGAVDATALTGDDRV  
 YGRVIATNGRGIYGPSPDEPPTPPQVDKGLVGLDNGDNRINISTDLTKMKRYILK  
 SIEDLPVEDDLWAADINGDGKINSTDYLYLKKYLQAIPELPKK"



QY 265 ProValPheTrpSerArgAspGlyGlyAlaThrTrpClnAlaValProGlyAlaProThr 284  
Db : : : : :  
1106 AGTATTTCACCGTACGACGGGGTGTCCACCTGGAAAGCAGTTCGCCGACCACTAAG 1165  
QY 285 GlyPheIleProHisIleValPheAspProValAsnHisValLeuTyrIleAlaThr 304  
Db : : : : :  
1166 GGAATCTCTCCACACCGGGTTTG---GCATCCACGGAATGTGTATATACTTAT 1222  
QY 305 SerAsnThrGlyGlyProTyrAspGlySer---SerGlyAspValTrpIlePheSerVal 323  
Db : : : : :  
1223 GGTGATACCTGCGTCTTATGACGCAACGGAAGGTCAGCTTTGGAAGTTCAATACA 1282  
QY 324 ThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyr 343  
Db : : : : :  
1283 CGTACAGGGGAATGGATAGATATACCCCGATACCTTATTCAGACAGTCAATTC 1339  
QY 344 PheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThr 363  
Db : : : : :  
1340 TTCTGCTTTGACGACTTGCAGTGGACAGCAAACTCTGCATTAATAATGGTAACCTTC 1399  
QY 364 GlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrp 383  
Db : : : : :  
1400 ATGAACGCGTGTGGCGGATGATATATTTTCGACGACTACGCGGAGCTACATGG 1459  
QY 384 ThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIle 403  
Db : : : : :  
1460 AGAATATCTGGGAATGGGAATGTATCTCTGACGTATATGCAATATGAAATAGATATT 1519  
QY 404 SerAlaGluProTrpLeuThrPheGlyValGlnProAsnProValProSerProIle 423  
Db : : : : :  
1520 TCGCAGCAGCGGTGGTGGATTTGGGGAACGTGAGAAACAGCTGCCGGAATCAATCCGAAA 1579  
QY 424 LeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyr 443  
Db : : : : :  
1580 CTGGATGATGATAGTGCATGATAGATGATGACCGTTTAATTCGACCCGATGATGAT 1639  
QY 444 GlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGln 463  
Db : : : : :  
1640 GTTACCGGTGCACTATCTATGTTGTGACAACTTCTACTGCTGGACAGAGCGGCAAA 1699  
QY 464 IleHisIleAlaProMetValIleGlyLeuGluGluThrAlaValAsnAspLeuIleSer 483  
Db : : : : :  
1700 GTAAATATCGAGTAAAGCTACCGGAATAGAGAATGTGCGGTATTTAGACCTGGTAAGC 1759  
QY 484 ProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAla 503  
Db : : : : :  
1760 CCGCCGAGGTGCACCGCTTGAAGTGCAGTGTGGCAGCTTGTGCGTTTGTTCATGAT 1819  
QY 504 AspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSer 523  
Db : : : : :  
1820 GACCTGAAAGTTGGTCCGAAAAAATGCACGTT---CCTTCTTATTTCTTCAGGTACGGGA 1876  
QY 524 ValAspTyrAlaGluLeuAsnPro-----SerIleIleValArgAlaGlySerPhe 540  
Db : : : : :  
1877 ATTGATTATCGGAGCTTGTTCGGAACCTTTATGCGATTGTTGCAAAAGCTGATTTGAT 1936  
QY 541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 560  
Db : : : : :  
1937 GATGTA-----AGAAAGATTCTTTCTTATGACGAGGGAAGGAAT 1978  
QY 561 TrpPheGln---GlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer 579  
Db : : : : :  
1979 TGGTTCACCCACTAATGAGCACCACAACTCGGTAGGCGCGGTTCGGTTGCGGTGCA 2038  
QY 580 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla 599  
Db : : : : :  
2039 GCCGATGCAAAATCAGTTATTGGACACCG---GMAAATGCAAGTCTCGAGTT---ACA 2092  
QY 600 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg 619  
Db : : : : :  
2093 ACGGACACGGAACCTCATGGAAGGTTGTGACAAATCTTGGTATGGGTGGGTGGCA 2152  
QY 620 SerAspArgValAsnProIlePheTyrAlaLeuSerAsnGlyThrPheTyrArgSer 639

Db 2153 TCCGACCGTGTGAACCGGTAAATAAATCTACGCAATTTCTATAACGGCAAAATCTATATAAGC 2212  
QY 640 ThrAspGlyGlyValThrPheGlnProVal---AlaAlaGlyLeuProSerSerGlyAla 658  
Db : : : : :  
2213 ACGGACCGTGGATTAACTTTACCGATACAAAGGACCGCAGCTTCCCAGATCGGTAAAC 2272  
QY 659 ValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAla--- 677  
Db : : : : :  
2273 -----AAGATAAAAGCGGTACCGGGCAAGGAGACATGATGCTTGTCTGCAAGA 2323  
QY 678 SerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyVal 697  
Db : : : : :  
2324 GAAGCGGATTTGTGGAGGTCCACTGACGCTGGATATACGTTTGAGAAACTTCCCAATGTT 2383  
QY 698 SerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaVal 717  
Db : : : : :  
2384 GACACAGCTCATGTGTAGGCTTCGGAAGGACGACCGGACAGGATTACATGGCGATT 2443  
QY 718 PheValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThr 737  
Db : : : : :  
2444 TACATTACCGGTAAATTTGACAATGTTTGTAGGATTTCTCCGTTCCGATGATCGCGCAAG 2503  
QY 738 ThrTrpValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThr 757  
Db : : : : :  
2504 ACATGGGTGCGTATCAACGACGACGACGATATGCGCTGTTCATCTACTGCAATAACA 2563  
QY 758 GlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyr 777  
Db : : : : :  
2564 GGTACCGGACGATATACGACGCTGATATATGTCACCAACGGAAGAGGATTTGTTTAC 2623  
QY 778 GlyAspIleGlyAlaProSerGlySerProSerProSer 791  
Db : : : : :  
2624 GGCAGAA-----CCTGCTTCAGATGACGCTGTACCCACT 2656  
RESULT 8  
ABI292929  
LOCUS 2579 bp mRNA linear PLN 08-AUG-2000  
DEFINITION Agaricus bisporus mRNA for CEL6 protein.  
ACCESSION AJ292929  
VERSION AJ292929.1 GI:9755896  
KEYWORDS cel6 gene.  
SOURCE Agaricus bisporus  
ORGANISM Agaricus bisporus  
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
Agaricales; Agaricaceae; Agaricus.  
REFERENCE 1  
AUTHORS Morales-Almora, P. and Thurston, C.F.  
TITLE Molecular analysis of the cellulolytic genes in Agaricus bisporus  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2579)  
AUTHORS Morales-Almora, P.  
TITLE Direct Submission  
JOURNAL Submitted (07-AUG-2000) Morales-Almora P., Microbiology, King's  
College London, 150 Stamford Street, London, SE1 8WA, UNITED  
KINGDOM  
FEATURES  
source Location/Qualifiers  
1..2579 /organism="Agaricus bisporus"  
/mol\_type="mRNA"  
/strain="D649"  
/db\_xref="taxon:5341"  
/clone="PMRT152"  
/dev\_stage="mycelium"  
/country="United Kingdom"  
1..2579 /gene="cel6"  
28..2448 /gene="cel6"  
/function="putative cellulase"  
/codon\_start=1  
/product="CEL6 protein"  
/protein\_id="CAC02964.1"





425 GGGNATATGCGCGTCCGGATGGGTGAGAGATTGGCGGTCCGACCGCAAGAAATAGC 484  
Db  
186 IleLeuTyrPheGlyAlaProSerGlyLeuTyrArgSerThrAspSerGlyAla 205  
Qy  
485 ATTCTCTACTTTGGAGCGCGAGTGGCGACCGCTTGTGAAGAGACACAGATTATGGAGCG 544  
Db  
206 ThrTrpSerGlnMetThrAsnProAspValGlyThrValIleAlaAsnProThrAsp 225  
Qy  
545 ACCTGGTCGAATGTCACCTCTTACGTGGAGTGAACAGTATTTCAGGATTTCAGGCTCG 604  
Db  
226 ThrThrGlyTyrGlnSerAspIleGlnGlyValIleValTrpValAlaPheAspLysSer 245  
Qy  
605 ACC-----TACACTTCGACCCCTGGGTGATTCTCTGGGTGACCTTTGATTCGACCTCT 658  
Db  
246 SerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnPro 265  
Qy  
659 GGTAGCTCGGGTCTGCGACCGCTAGGATCTTCGTTGGAGTCGGGATGCTGGCAAGTCG 718  
Db  
266 ValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThr--- 284  
Qy  
719 GTCTTCAAGTCGGAGGATCGCGGTGCTACTTGGCTTGGGTGTCTGGAGAGCCTCAGTAT 778  
Db  
285 GlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThr 304  
Qy  
779 GGTCTTCTGCGGCACAGGGGGTCTTGTCCCTGAGGAGAGACGCTGTATCATCTCTAC 838  
Db  
305 SerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThr 324  
Qy  
839 CGGNATGGCGCTGCTCTACGATGGGACAAACGGAACAGTCCCAAGTACACATCACC 898  
Db  
325 SerGlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPhe 344  
Qy  
899 AGCGGATGTGGACGGATATAAGTCCC-----ACCTCCCTTGGCGAGTACCTACTAC 949  
Db  
345 GlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGln 364  
Qy  
950 GGCTACGGTGTCTGCTCGGTGACCTCAAGTTCCTCGGACCTTGTATGGTGGCTGCAC 1009  
Db  
365 IleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThr 384  
Qy  
1010 AACTGCTGGTGGCGGATGACTCATCTCCGTAGCACTGACTCTGGAGCGACCTGGTCC 1069  
Db  
385 ArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSer 404  
Qy  
1070 CCTATTGGAGTGAACCGTTCATCCGAGCATCAATTACTACTACGCTATGATATCAGC 1129  
Db  
405 AlaGluProTrpLeuThrPheGlyValGlnProAsnPro---ProValProSerProLys 423  
Qy  
1130 AACGCCCTTGGATCCAGGACACGACTTCGACCGACCAATTCGCGGT-----CGC 1180  
Db  
424 LeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyr 443  
Qy  
1181 GTGCGTGGATGTTGAAGCCCTTGGCATTTGACCCCATTTGACTCCAACTGCTCTAC 1240  
Db  
444 GlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGln 463  
Qy  
1241 GGACGGGCTGACGATATGCGGACACGACTGACCAATTGGGACTGGAAGCACAAT 1300  
Db  
464 IleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSer 483  
Qy  
1301 GTGACGTCGAAGTCACTGGCGGTGGTATTGAGGAGATGCGTCTCGGACTGATCAGC 1360  
Db  
484 ProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAla 503  
Qy  
1361 CCGCCCGGGGACCGGCTGCTGCTCGTGTGGGACGACGCGCGGTCTTACCACATCG 1420  
Db  
504 AspValThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSer 523  
Qy  
1421 GATCTGAGCAGCGCCAAACCAAGGTACCAACCGCGGACCTACGGCACAACCAAGCGC 1480  
Db  
524 ValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSer 543  
Qy  
1481 ATCGACTACCGCGCAACAGCCTTCCACATGTCGCGGTCCGAGCTACCGAC----- 1534  
Db

544 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln 563  
Qy  
1535 GACTATCCGACG-----CTGCTCTGTCTCGAATCTCGATCGACTTGGTATGCA 1585  
Db  
564 GlySerGluProGlyGlyValThrGlyGlyThrValAlaSerAlaAspGlySer 583  
Qy  
1586 GACTACGGCGGTCAACGACGACCGGCGGTGGCTCTTCCCGGATGGTGAC 1645  
Db  
584 ArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGly 603  
Qy  
1646 ACGGTGCTTCTCATCTCCAGCACCTCGGGCGGTGCTGTCGAGTCGACGGC----- 1699  
Db  
604 AsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal 623  
Qy  
1700 --ACCTTACTGCGGTTCCTCTCTCCGTCGCGGCGGTCTATCGCTCCGACAAAGTCC 1756  
Db  
624 AsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGly 643  
Qy  
1757 GACAAACAGGCTTCTATGTGGATCCGCGGCGCAATCTACGTGTGGAAGACACAGCG 1816  
Db  
644 ValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPhe 663  
Qy  
1817 ACCTCTTCCACAGACCGGTGAGC---CTGGGTCCAGCACCGCTCAACGCCATC--- 1870  
Db  
664 HisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHis 683  
Qy  
1871 CGCGGCATCTCTCCATCGCGCGATGCTCTGGCGGTCCAGCACAAAGGCGCTCTGGCAT 1930  
Db  
684 SerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 702  
Qy  
1931 TCCACGGATTACGGTAGTACGTTTCAACGAGAGAGAGGAGGACCACTGGCAGGTGATC 1990  
Db  
703 ValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThr 722  
Qy  
1991 TTTGGTGTGGCAAGCGCTCTCCACGGCGACTACGCTGCTATCTACGGTCTTTCACC 2050  
Db  
723 IleGlyGlyValThrGlyValAlaTyrArgSerAspCysGlyThrTrpValLeuIle 742  
Qy  
2051 ATCAGCGTGGCGGGCTGTTCAAGAGCGAGAGAGGAGGACCACTGGCAGGTGATC 2110  
Db  
743 AsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAla-----IleThrGlyAspHis 760  
Qy  
2111 TCGGATCGTGCACGAGTTCGGCTCG---GGTCCGCCAACGTTGGTGAACGCGACCTG 2167  
Db  
761 AlaAsnLeuArgArgVal-TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780  
Qy  
2168 CAGACCTACGGCGCGGTGTTTCTGGGCAACGACGCGCGGCGCATCTTCTACGGCAAT-- 2225  
Db  
780 eGlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLe 800  
Qy  
2226 -----CCGACGGGAGCTCCCGGCGCCACGCGGCGGACCGGACGCGGACGAC 2275  
Db  
800 uSerProSerProSerProSerProSerProSerProSerProSerProSerProSerPr 820  
Qy  
2276 ACCTTCAAGACTCG-ACGACGGTATCCAGCAGTTGAAACCAACCAACGAGCGGCTC 2334  
Db  
820 oSerSerProSerProSerProSerProSerProSerProSerProSerProSerPr 840  
Qy  
2335 GACCACGCTTCCAGCAACACCGTGAAGACCAACGACTAGTCTGTCAGCAGCAGCTTCAA 2394  
Db  
840 oSerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPr 860  
Qy  
2395 GGCCAGCTTCGACCAACCAACAGACTACCAACCAACCAACCAACCAACCAACCAACCA 2454  
Db  
860 oSerProThrProSerProSerProValSerGlyGly 872  
Qy  
2455 CACCGGACGCGCTCGGCTAGCTCAGTGTGGCGGC 2491  
Db

RESULT 10

AY040839

LOCUS

DEFINITION

3959 bp DNA linear PLN 01-APR-2002  
Aspergillus niger endoglucanase C (eglC) gene, complete cds.

ACCESSION	AY040839	Db	1077	-----CTGGACTCGGTCCACGCTGCCGCATCGCAGGCATATACCTGGAGAAATGTG	1127
VERSION	AY040839.1	Qy	58	AlaIle	60
KEYWORDS		Db	1128	GTAC-AGGGGGTATGACATATCTTCAATATAGGGGATTGACAACTAAATGATCAGCG	1186
SOURCE	Aspergillus niger	Qy	61	GlyGlyGlyPheValAspGlyValPheAsnGluGlyAlaProGlyIleLeuTyrVal	80
ORGANISM	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.	Db	1187	GGCGAGGCTTACGCGGGTATCGTCTTCAACCTTCAGCTAAAGGTGTCGATATGCT	1246
REFERENCE	1 (bases 1 to 3959)	Qy	81	ArgThrAspIleGlyGlyMetTyrArgTyrAspAlaAlaAsnGlyArgTyrIleProLeu	100
AUTHORS	Hasper,A.A., Dekkers,E., van Mil,M., van de Vondervoort,P.J. and de Graaff,L.H.	Db	1247	CGACACGACATCGCGGTGCATATCGGTGAAC---TCGGACGATACATGGACCCCGT	1303
TITLE	EglC, a new endoglucanase from Aspergillus niger with major activity towards xyloglucan	Qy	101	LeuAspTyrValGlyTyrAsnAsnTyr-----GlyTyr-Asn-----	112
JOURNAL	Appl. Environ. Microbiol. 68 (4), 1556-1560 (2002)	Db	1304	ATGGACTGGGCAACAACTCTAACTGGTACGTAAGCTTCTCTACGAGATGCTAGGCTC	1363
MEDLINE	21914059	Qy	113	-----GlyValValSerIleAlaAlaAspProIleAsnThrAsnly	126
PUBMED	11916668	Db	1364	ACTGTCAGGCACGATTGGGGCATTTGATCGCATCGCACCCCGCTCGATACCGATCG	1423
REFERENCE	2 (bases 1 to 3959)	Qy	126	sValTyrAlaAlaValGlyMetTyrThrAsnSerTyrAspProAsnAspGlyValAlaIle	146
AUTHORS	Hasper,A.A., Dekkers,E. and de Graaff,L.H.	Db	1424	GGTATATGTTGCAGTTGGCATGTACACCAACGACTGGGATCCCAATGACGGTCCATCT	1483
TITLE	Submitted (18-JUN-2001) Molecular Genetics of Industrial Micro-organisms, Wageningen University, Dreijenlaan 2, Wageningen 6703HA, The Netherlands	Qy	146	uArgSerSerAspGlnGlyAlaThrTyrGlnIleThrProLeuProPheLeuGlyG	166
JOURNAL	Location/Qualifiers	Db	1484	CGCGTCGACCGACCGAGCGGCGATACCTGGGAAGAGACAAAGCTTCCCTTCAAAGTTGGAG	1543
FEATURES	1. .3959	Qy	166	YAsnMetProGlyArgGlyMetGlyGlu-----	175
source	/organism="Aspergillus niger"	Db	1544	GAATATCGCGTCCGGGTGTTGGCGAGGTAACTATATCTTGGGTTTGTGTAGCATAC	1603
gene	/mol_type="genomic DNA"	Qy	176	-----ArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyA	191
mRNA	/db_xref="taxon:5061"	Db	1604	ACTGACCGTGCAGCGTCTGGGTGATCGATCGATGACCAACAGCATCTGTACTTTGGTG	1663
	<1038. .>3841	Qy	191	laProSerGlyLeuGlyLeuTyrArgSerThrAspSerGlyValaThrTyrSerGlnMet	211
	/gene="eglC"	Db	1664	CCCGAGTGGCAATGGCTTGTGGAAAGACACGAGTACGGCGAGACATGGTCCAATGCA	1723
	join(<1038. .1137,1185. .1329,1373. .1571,1618. .1858,	Qy	211	hrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln	231
	1908. .1961,2007. .>3841)	Db	1724	CTGCCTTCAAGTGGACCGCGCACTTATTTCCAGATTCAGGCTCCGACC-----TACACGT	1777
	/product="endoglucanase C"	Qy	231	erAspIleGlnGlyValValTyrValAlaPheAspLysSerSerSerSerLeuGlyGlna	251
	/gene="eglC"	Db	1778	CTGATCCGGTGGGAATTCGCTGGGTACCTTCGACTCGACCTCAGGGTCCAGTGGCTCTC	1837
	/codon_start=1	Qy	251	laSerLysThrIlePheVal-----	257
	/product="endoglucanase C"	Db	1838	CGACTCCGGCATCTTTGTCCGTATGTCAAAAATTTAGATATCTCAGACAGCCACCTG	1897
	/protein_id="AAK77227.1"	Qy	258	-----GlyValAlaAspProAsnAsnProValPheTyrSerArgAspGlyGly----	273
	/db_xref="GI:19879406"	Db	1898	ACCATTACAGGTGTGGTGACACTGGAGAAATCCGTTTGTGTCCGAGATGCTGGCGAA	1957
	/translation="MRTFAALVAGYLLDSVHAASQAVTWKVVVTGGGGTTPGIVP	Qy	274	-----AlaThrTyrGln	277
	NPSAKVARTDIGAYRLNSDWTPLMDNNSNWHWDGIDALATDPDTRVY	Db	1958	ACTTGTAAAGCTTCTCCAGATTCGCAATTAACCCCATCTAACCACTTGGGACTTGG----	2014
	AVGMYNDDPDGSLRSDQDQDWEETKLPKGVGNMFGVGRGLAVDPDNLIL	Qy	278	AlaValProGlyAlaProThr---GlyPheIleProHisLysGlyValPheAspProVal	296
	YFGASGLKWDSTDGTEVSNVAFKWTGYTFQSSSTYSDFVGLAWTFDSTGS	Db	2015	---GTAICTGGAGACCCATGTACGGCTTCTCCCCCAATAAGGCATCTCTGCGCTTCT	2071
	SGSPTRIFVGVDTGESVFSADAGETWTWSEPMYGLPHKGLILSPSEHTLYISY	Qy	297	AsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerGly	316
	SNAGDITNGTGVHKYNTISGVMTDISPTSMTDYGYGLAVDLQVPGTVMVAAL	Db	2072	GAGCACACACTCTACATTTCTTACTCCACCGCGCGGTGCCATGATGGCACCRAATGGC	2131
	CWPDLEIWRSTDGCTWSPINAWGVPSINYYSYDINSAPWLDDTSTDPEPVRVG	Qy	317	AspValTyrLysPheSerValThrSerGlyThrThrArgIleSerProValProSer	336
	WVVEALADPPSDHMLYGTETIYGHDLQWDSHNVTIESLAVGIEMAVLGLIT	Db	2132	ACCCTCCCAAGTACACATCACCAGCGCGCTCTGGACCGGACATCAGTCCACCTCCATG	2191
	PPGPAALLSAGVDGDFYHLSLTPAQYTHFTPTYSTNGIDYAGNKPANIVRSQSD				
	SDPTLALSSFGESVDAVAASSTATQVALSADADTILLMNSDGYARSANATLPS				
	VSSLPSGAVIASDKANNTPYFAGSGSFYLSDDTAATFTVTTTLCGSTTANALRAQPS				
	LAGDVWVSTDGLPHSTNYGKSPQISGCTGWSFGFKPSDGDYFVLPFGFTVDG				
	VTGLFKTEQGVNWOIISDAEHGFGSASNVNVDLQNYGRVFGVGTNGRIGIPLYDPSG				
	TIPLSATASSASTAVKSTSTSTSKVSSSTVSSSTATTITTSIKSTLTITTKS				
	SSSTSTSTSTATASAYGCGGSGGTGFTQCPGSGWTCTYENETYSQCKSLPIGIATR				
	G"				

## ORIGIN

Alignment Scores:  
 Pred. No.: 5,08e-43 Length: 3959  
 Score: 1690.00 Matches: 390  
 Percent Similarity: 53.46% Conservative: 128  
 Best Local Similarity: 40.25% Mismatches: 315  
 Query Match: 32.91% Indels: 137  
 DB: 8 Gaps: 26

US-09-917-376-1 (1-957) x AY040839 (1-3959)

Qy	18	SerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProIle	37
Db	1026	TACTGTCACCATGGCGGTAGATTCGACCGCTCGTAGCG---GGCTACCTG-----	1076
Qy	38	AlaIleThrAlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTyrSerAsnVal	57

















Db 161366 GCCGTATCGGGGACCGGCTTCGTACACCGGCTGCTCTTCACCCCTCGTACGGCGT 161307  
Qy 77 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArg 96  
Db 161306 CTCGCTAGCCCGGACCGATCGGGCGCTACCTACCGCTGGAGCACCAGCGGCGCCCGC 161247  
Qy 97 TrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSer 116  
Db 161246 TGGACCCGCTCATCGACCACTCGGCTGGAGCACTGGAACCTCTCTCGGCGTCGAGCG 161187  
Qy 117 IleAlaAlaAspProIleAsnThrAsnLeuValTrpAlaAlaValGlyMetTyrThrAsn 136  
Db 161186 ATGGCGTGCACCCACCGACCGGACCGGCTTACCTCGCGTGGGCACCTACGCCAG 161127  
Qy 137 SerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGln 156  
Db 161126 TCGTGGCGGGCAAC---GGCGGGTCTCGCTCCGAGACCGCGGCGCACCTTGACC 161070  
Qy 157 IleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArg 176  
Db 161069 CGCACCGACCTGACCGTGAAGCTCGGGCAACGAGGACGGCGCGCGGTGAGCGA 161010  
Qy 177 LeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 196  
Db 161009 CTCCTGTCACCGCGCGACGACGACACCTCTGCTGGGACG---CGGCACGACGG 160953  
Qy 197 LeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVal 216  
Db 160952 CTGCTAAGTCGACCGACCGGGCGGCACCTGGGACGCGCGGCTTCCCG--- 160899  
Qy 217 GlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVal 236  
Db 160898 -----GCGAAGCGCACTCTCCGGCGAGGATC 160869  
Qy 237 ValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerIleThrPhe 256  
Db 160868 GTGTTC-----CTGTCGCGCGGGCGCACCGCTCAC 160836  
Qy 257 ValGlyValAlaAspProAsnAsnPro-----ValPheTrpSerArgAsp 271  
Db 160835 GCGCGGTGGGGTACGCGGACGCGACCTCGGGCACGCGCAACCTGTACCGACGCGCGAC 160776  
Qy 272 GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe-----IlePro 288  
Db 160775 ---GGACAGACCTGGGGCGCGTCCCGCGCGCTTCCGCGACCTCCGCAAGTCCCG 160719  
Qy 289 HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 308  
Db 160718 CTCGCGCGCGTACGACACGACACCGCGAGCTGTACGTGACGTACGCGGACGCCACC 160659  
Qy 309 GlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrp 328  
Db 160658 GGGCGCGGGCGCGTCCGACGCGCAGCGTGCAAGCTGCGTACCGCGCACGCGGACGTGG 160599  
Qy 329 ThrArgIleSerProVal-----ProSerThrAspThrAlaAsnAspTyrPhe 344  
Db 160598 ACCAGTCAACCCCGTGAAGCGGGGAGACCGGCGGACCGGCTCGCGGACACCTTC 160539  
Qy 345 GlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGln 364  
Db 160538 GCCTACGGGGGTGCGCGTCGACGCGCGCGCGCGCGCACCTCTGCTGCTCCACCAAC 160479  
Qy 365 IleSerTrpTrpProAspThrIlePheArgSerThrAspGlyAlaThrTrpThr 384  
Db 160478 AACCGTGGGGCGGACGCGTCTTCGGTCCACGCGCGCGCGCTACCTGACG 160419  
Qy 385 ArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSer 404  
Db 160418 TCCTCAAGAC-----GCCCGCGTTCGACGCTGC 160386  
Qy 405 AlaGluProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeu 424  
Db 160385 GAGACTCCCTCTCTCGACTGGGCGGACGAC-----AAGCGGAGTTC 160344

Qy 425 GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGly 444  
Db 160343 GGCTGGTGGATCCAGGCGCTCGCGTCCAGCCGTACGACTCCAGCACGTCGTGTACGG 160284  
Qy 445 ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlnIle 464  
Db 160283 ACCGGCGGACCTCTACGGCACCGCGACCTCAAGCGTGG----- 160242  
Qy 465 HisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerPro 484  
Db 160241 -----GACCGCGGATCCGCGCTTGAGGAGCGCGTCCGCAACTGATCTCGCC 160188  
Qy 485 ProSerGly---AlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAla 503  
Db 160187 CCGGTGGGAGGACACCTGATCAGCGGACTCGGGGACATCGGTGTGTATACCAAGAG 160128  
Qy 504 AspValThrAlaValProSerThr---IlePheThrSerProValPheThrThrGlyThr 522  
Db 160127 CGGCTCACGGGCTCTCGTCGCGGCGATGCGACGAAACCCGCTGTCGGTTCGGGACG 160068  
Qy 523 SerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspPro 542  
Db 160067 GGACTCGCGGACGCGCGCGGCGGCGGCGGATGTCGTCGCGACCGGCTGGGGCGACAC 160008  
Qy 543 SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp--- 561  
Db 160007 GGCACCGC-----CGTACTCCACGACGCGGCGGCGGCGGCGG 159966  
Qy 562 ---PheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla 580  
Db 159965 CCCCTCGAGGCGGCGGCGGACATCGCAAGGACGCGCGGACCGATCGCCACCACTGCC 159906  
Qy 581 AspGlySerArgPheValTrpAla-----ProGlyAspProGlyGlnProVal 596  
Db 159905 GACGCGCGCACACTGCTGTGCTCTCTGTCACCTGGGACGGC-----ACG 159861  
Qy 597 ValTyrAla-----ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyVal 612  
Db 159860 ACGTACGCGGCGGCGGCTCGACGCGACACCGCGGCGGCTGTCTCGAGGTCTCTCTTC 159801  
Qy 613 ProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla----- 630  
Db 159800 CCGAAGGCGGCGCACCGGTCGCGGCGGCGGCGGCGGCTTCTACGCTACGAC 159741  
Qy 631 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla 650  
Db 159740 TTCGACAAATGGAACGCTATACCGCGACACTGACAGTGGCGGCTTCTCACGCGCGTGG 159681  
Qy 651 AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGlu 670  
Db 159680 GCGCGACTGCCC---TCCGCGCAGCGCGGCTCAAGCTGGTCCGCGCGCGCGGACGAGC 159624  
Qy 671 GlyAspLeuTrpLeuAlaAlaSer---SerGlyLeuTyrHisSerThrAsnGlyGlySer 689  
Db 159623 GCGGACTGTGGCTCTCGCCAAATGGAACCGGCTCTACCGGTCACCGCGCGGCGGAC 159564  
Qy 690 SerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAla 709  
Db 159563 ACCTTCGCCAGATCGACAGCTGTGGGCGCTCGTACACCTCTCGGCTTCGGCAAGCGGCGC 159504  
Qy 710 ProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla 729  
Db 159503 GACGCGCGCGACTACCGCGCGATCTACAGGTGCGCTCGCGGAGACCATCCCGCGCTC 159444  
Qy 730 TyrArgSerAspAspCysGlyThrTrpValLeuIleAsnAspGlnHisGlnTyr 749  
Db 159443 TACCGCTCCGACGCGCGGCGGCGGACATGGGTCCGGATCAACGACGACGCCACCACTGG 159384  
Qy 750 GlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGly 769  
Db 159383 GGGTGGATCGCGAGGCGGCTCGTGGTGACCGCGCGATCCCGCGGCTCTACCTCGCC 159324

```
Qy 770 ThrAsnGlyArgGlyIleValTyrGlyAspIle-----GlyGlyAlaPro 784
Db 159323 ACCAACGGCGCGCATCAGTACGAGGAGCGCGTCTGTGATGCCGGGCTCACGGATGCCA 159264
Qy 785 SerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerPro 804
Db 159263 CCGGGGCGCGCATCTCTGTTGGTGGCG---ACTACACCCCGAGCAGTGGCCCGAGGAGG 159207
Qy 805 Ser------ProSerSerSerProSerPro----- 812
Db 159206 TCTGGCAGGAGGACGTACGCCTCATGAGGAGCGCGCGTCACTCGTCCACCGTCGGCG 159147
Qy 813 ---SerProSerProSerSerPro----- 820
Db 159146 TCTTCTCTCTGGCGGAGCTCGAACCACCCCGGGCGGAGGAGTTCGGTGGCTCGACCC 159087
Qy 821 ---SerSerProSerProSerPro----- 828
Db 159086 GGTCTATGACCTGATGCAGAGAACGGCATCGGCTGCTGCTGGCCACCCACCGCCT 159027
Qy 829 SerProSerPro----- 832
Db 159026 CGCCCCCGCTGGATGGGGCGCTGCACCCCGAGACCTGCCCGCGAGGAGCGGCC 158967
Qy 833 ---SerProSerArgSerProSerProSerAla 842
Db 158966 GCATCGAGTGGTGGCGCGCGCCAGCACTTCGCCCACTCCAGTGGCGCTCTACCGCGCT 158907
Qy 843 SerProSerProSerSerProSerProSerSerSer-----ProSerSer 858
Db 158906 ACCCGCGCGCATCACCGAGGACCTCGCCCGCGCTACCGCGGCGCATCCCGCGCTCACGA 158847
Qy 859 SerProSerProThrProSerSerProValSerGlyValSer----- 874
Db 158846 TGTGGCACATCACACGAGTACTGCACTACGACTGGCGGAGCGGCGCCACCGCCT 158787
Qy 875 -----ValGlnTyrIleAsnAsnAspSerAlaProGlyAsp 886
Db 158786 TCCCGCGCTGGCTCCAGGCGAGGTACGGCAGCTCGACGCACTCAACACGCGCTGG--- 158730
Qy 887 AsnGlnIleAspPro---GlyLeuGlnValValAsnThrGlySerSerSerValAspLeu 905
Db 158729 ---GCACGGCGCTTCTGGAGCGAGGCTACGACGAGTGGGAGGAGGATCTCCCGCGCG 158676
Qy 906 SerThrValThrValArgTyrTyrPheThrArgAspGlyGlySerSerThrLeuValTyr 925
Db 158675 GCCACGCCCTACATGAAACACCCACCGAGGTGCTGCACTTCAAGCGCTTCACTCCG 158616
Qy 926 AsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 945
Db 158615 ACGCACTCATGGAGTGT---TCGTCCGCAAGCTGACATGCTG----- 158574
Qy 946 ProAlaThrPro 949
Db 158573 CCGGCGCACACCC 158562
RESULT 14
LOCUS AE011809 14520 bp DNA linear BCT 29-MAY-2002
DEFINITION Xanthomonas axonopodis pv. citri str. 306, section 187 of 469 of
the complete genome.
ACCESSION AE011809 AE008923
VERSION AE011809.1 GI:21107967
KEYWORDS Xanthomonas axonopodis pv. citri str. 306
SOURCE Xanthomonas axonopodis pv. citri str. 306
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
Xanthomonadaceae; Xanthomonas.
REFERENCE 1 (bases 1 to 14520)
AUTHORS da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
Chambergo,F., Ciapina,L.P., Ciccarelli,R.M.B., Coutinho,L.L.,
Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira,Jr.,H.A.,
Rosi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
Taki,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
Kitajima,J.P.
Comparison of the genomes of two Xanthomonas pathogens with
differing host specificities
Nature 417 (6887), 459-463 (2002)
22022145
12024217
2 (bases 1 to 14520)
da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
Chambergo,F., Ciapina,L.P., Ciccarelli,R.M.B., Coutinho,L.L.,
Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira,Jr.,H.A.,
Rosi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
Taki,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
Kitajima,J.P.
Direct Submission
Submitted (28-Nov-2001) Departamento de Bioquimica, Universidade de
Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
Brazil
FEATURES
source
1. 14520
/organism="Xanthomonas axonopodis pv. citri str. 306"
/mol_type="genomic DNA"
/strain="306"
/db_xref="taxon:190486"
/note="pathovar: citri"
87. 2342
/gene="celA"
/note="synonym: XAC1770"
87. 2342
/gene="celA"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/trans_table=11
/product="Cellulase"
/protein_id="AA036634.1"
/db_xref="GI:21107968"
/translaton="MIAVIDPTKRFSAHAWLTAVCLTVLLFWTSEASADEPCTP
DPYQWHSVACGGGFTVGLFHPAERDLAYARTDVGGAYRWDAQOQWALDMLGAD
DNLNGIDAFADVPADPNALYLAAGTMYHERAGNAVLRSFDRKQFQERADLPFLGG
NQLGRANGERLAVDPHGRVLLGSRDGRGAWRVEGFPFADALAGATARN
HVGSEQAVGIAFVVDYASGAGAAATPRIYGVSTAQTSYVSEDAQRTWSAVAGQPK
GLRPSHMVGNAGWYLSYGRPGPDLMAAGLWYDATQGRWRISPTIPQATGDF
GWGAVADVPQHPVLLASTFRRTPRDEVSFGDGRSWVPLAAQPHSAAPTAH
ATPHWIGALADFPDSNHATFTVGTIWSRNLTQTSFQQLQWFWQDRGLSETVPLD
LLSPMAGHLLSALGIDFHRDALTQALQVLPRLTNGESIDGAGQAPQWVRSGT
VDRNRNEIRALYSQDGTGWAFSEPRGQAGTIAIADASQVWVVDGQGWRT
GDFGRWQVORGLPDTAVVADRVDAQRWADRVSGLYSDSGASGASGQVQVGS
PADERARPOLRPPDPRWAGVYLASPTLGVMRWONGOLRTLSKSEARSLGKALRA
GAPPALYLAGRVAGVDFGIFRSDDDEGAHWRINDDAHRFGKPSYVTGDPRIAGRVFAT
GGRGIFGDPDR"
2431. 4347
/gene="XAC1771"
```











```
gene
4555. .6747
/notes="c1rA"
/genes="XCC1750"
CDS
4555. .6747
/genes="c1rA"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/codon_start=1
/transl_table=11
/product="TonB-dependent receptor"
/protein_id="AAM41041.1"
/db_xref="GI:21112843"
/translation="MLPPLGLKPCWDYHASTSPRLPRPDVAATTGGQVRAPAROK
IESTCVLGSTWSVQHRHLPAGRAVQGRSIFRFRSVMASVATLLSAQYADATTT
AAPASDATQQLDVTQVTTGRSSVTKAQLVQNAEIVDSIVAEDIGKLPDNNVAEL
ORISGIGSNYGESSIAIRGLTVRTELNGRDIPTANDGRGLSPEDVSAELGGVN
VYKNSADMIEGLGTVDLRTELPEYGRKTAGSVQYNYVDLADGKPAFGLFSN
RWQGTGEGTILVNYQQSQPPRODTISIEPWTQTLDPGYEQGVSVPHGAGINTI
GRERKIGAFAMQRNPDSVEVYTVQLRSDYDSFQWDSYFFALTQRNPQGLPGIQN
DRNEFYGFQNPVTESNTLSRHSVTTDYSLGAKWTYNEKLTLSLTDYQYVDATTG
TRYIVATGQDTSQFNVDFRGLPRLSDTASGAGYLDVNNYDGRWYHLDNDKNK
GTEFAWADMIDAFSDVFRSPKAGVYTRDRAETKGNTRYFMCLNTACGAATSPFA
APSGIIVNPIPIIDFRGSSSTFGRILTAAANSVSNYEQTATFEGASPLVFAENLNA
QREKTYAAGVLRFVDSVSDPFDGNIGVVRVTEVSGNVRGTGDAEGGLIPVAQ
QTYTDVPLSLNLRWMLSDQLWRFAASRGISRPTFDRLNPNLSLSTGT"
6853. .7455
/genes="XCC1751"
/notes="XCC1751"
6853. .7455
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/codon_start=1
/transl_table=11
/product="TonB-dependent receptor"
/protein_id="AAM41042.1"
/db_xref="GI:21112844"
/translation="MMYGLTFYKKVEGFIANVFNVEYDQVWQITRPVNGDAGKING
AELGYTFDFPLFGVGLQNTYTVDSSEAPSTATDNGQSLTVPLGLESKNSYN
AIIYSYETARFQGVAVNWRSDMLVTTAGNCTGNLVPVNGFGQLDASLRNINQVSI
SLDGNLLDTRRSYLGTSRYRDFVINDRRYGLTLRASL"
7567. .9822
/genes="XCC1752"
7567. .9822
/genes="XCC1752"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/codon_start=1
/transl_table=11
/product="cellulase"
/protein_id="AAM41043.1"
/db_xref="GI:21112845"
/translation="MIAMIDPPTKRRTTTGAWLAVLSLLLLLFSTPNVRAEPAISG
PYQWRSVAGGGGFTVGLVPHFAERGLAYARTDVGGAYRWDAQQWALTDLWIGADD
WNLMDGDAVADPADADALYLAAGTYMHERAGNAVLRFSNRTFERADLPFKLGN
QLRANGERLAVDPHGRVLLGLSRDAGLWRSDDRGARHAKVASFPDAALAGATARNH
VQREAOVGFATVYVDAAGNTGPTPRIYVGVSTEOTSLVKSVEDAGRSWAPVAGOPRG
LRPSHWAGSDQHWLSYGDOPGCDLMAGCALMKFTPAQGRWEI SPIQOPASGDGFG
WGA VADPOQPOGHWLSTFRRTPRDELKSVGCKHWAPLLADA VFDHSAARWTHA
THWMGALAIDPFDGNHALFTVGTGIWASRNLDQFAOPRPLQWFFDRGLEETPLD
LUSPMAGHLLSALGIDGFRHDELDRALQYAGPRLTNGESIDAAGQAPQWYVRSQT
VDRRNNEIRALYSRGGQKWTAFAPSEPPAGQAGSIATIGADAQVVMVAPERGWNRT
SDPQACQWRDGLPDTAVMADRVARVYADVASGOLYESTDARSFRATGVGVGS
PARDETRQLRDPHRAQVYLYASPCKGVMRWDDGTQLVLSQPDARSISLIGKALRA
GAPPALYLAGRVGVDGVRSDDDGGVQWQNRINDAHRFGRPISVTGDPRIAGRVYFAT
GGRGIFYGDP"
9926. .11839
/genes="XCC1753"
9926. .11839
/genes="XCC1753"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/codon_start=1
/transl_table=11
/product="sialic acid-specific 9-O-acetyltransferase"
/protein_id="AAM41044.1"
/db_xref="GI:21112846"
/translation="MTPSHISIRACLLGLVLLSSGSAQAQPSLPLFADGAVLQDRAMP
VWQATPGASITVTSFDGKQATKTAQDQWQKVLPAHTAGGPPYVLSVHGDDGALQVRD
VLVDVWLASQSNMELPAQASDQPOAANAANDQLRHFVKVPSKWSVQPETKLNGGD
WKAAMPENAGAFSNGVYFFAKELRASPVPIGIVDSITWGSALIEAMDAASLGLXAEQ
NQAIEAIKORDAAQAATKRIGRWPKVDGDTTQWRTPEDFDDWDSDSIPTVQQWELS
GYEGMDGIAYRITFTLTAABAKAGVQGVQIDSDSIYVNGQVQGVQTEKQNNLPRV
YAVPAALRAGVQIIVRVEDLSGGGGMHGPDEQRFVQIGSTQALAGKFRPAQVRV
SLTDNKNOLPITVNMHPLQPPFVKGIWYQGETNATDTKAVKRDQFAMIQOWR
AERGAELPLWQLANFKAGGDTGDLSPWALLRESQSTLALPATGQAVLIDIGNPT
DIPTNKRQDVHRLALAAHVAHYGETLYVSPAFARASYADQASVQDQLQGSALQVR
EGGTLQGFAGVAGADKRFHPAQARIVRSDA VAAVAVRVYGSNPADANLVNR
EQLFVSPFRTDTW"
ORIGIN
Alignment Scores:
Pred. No.: 7,12e-27 Length: 11910
Score: 1170.50 Matches: 299
Percent Similarity: 48.61% Conservative: 138
Best Local Similarity: 33.26% Mismatches: 350
Query Match: 22.79% Indels: 113
DB: 1 Gaps: 28
US-09-917-376-1 (1-957) x AE012276 (1-11910)
Qy 23 ThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSer 42
Db 7606 ACCAACACGAGCAGTCGTGGCTGCTTTGAGCCCTGCTG---CTGTTGCTCTTTTCGACG 7662
Qy 43 Pro-----AlaHisAlaAlaThrGlnProTyrThrTrpSerAsnValAla 58
Db 7663 CCCAACGTCGTGGCGCGAGCCGCCGCGCTACCACTGGCGCAGTGTGCC 7722
Qy 59 IleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeu 78
Db 7723 ATTGGCGTGGCGGATTTGTCCCGGTGTGCTTTCATCCGCGCGAACGTGCTCGGCC 7782
Qy 79 TyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIle 98
Db 7783 TATCGCGCACCGATGTGGGTGGCGGTACCGCTGGATGCCGCGCAGCGCAGTGGACC 7842
Qy 99 ProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAla 118
Db 7843 GCGCTGACGACCTGTTGGGCGCGCAGCAGCTGGAACCTGATGGGCATCGACCATTCGCC 7902
Qy 119 AlaAspProIleAsnThrAsnLysValTrpAlaValGlyMetTyrThrAsnSerTrp 138
Db 7903 GTCACCCCGCGATGCGATGCGTGTATCTGCGCGCGCCACCTATATGACGAAACGC 7962
Qy 139 AspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThr 158
Db 7963 GCGCGCAAT---GCCGCGAGTCTGCGCTGTTTCAACCGCGCGCCGACGTTCCGCGCGCC 8019
Qy 159 ProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAla 178
Db 8020 GACCTGCGCTTCAACTGGGTGTAAACACGCTGGCGCGCCCAATGGCGAGCGGTGGCG 8079
Qy 179 ValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrp 198
Db 8080 GTGACCCACACGATGGCGCGGTGCTGCTGCTGGGCTCGCGCGATGCC---GCCCTGTGG 8136
Qy 199 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 218
Db 8137 CGTAGCAGCATCGCGCGCGCAGCTGGCGGAAAGTGGCTGCTTTCCGACCGCGCTG 8196
Qy 219 TyrIleAlaAsnProThrAspThrGlyTyrGlnSerAspIleGlnGlyValValTrp 238
Db 8197 GCGCGTCCACCGCAGCAATCATGTTGGGCGCGAGCGCGGTG---GGGATCGCTTT 8253
Qy 239 ValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrPheValGly 258
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 14, 2004, 08:53:04 ; Search time 848 Seconds  
(without alignments)  
4794.252 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MDRSENIRLTMRSRLVSL.....RASFGVNPATPTADTYLQX 957

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2\_1/USPTO.spool/US09917376/runat\_14052004\_085256\_4609/app\_query.fasta\_1.1095  
-DB=N Geneseq 29Jan04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376.@CNG\_1\_1\_581.@runat\_14052004\_085256\_4609 -NCPUP=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq 29Jan04:.\*  
1: Geneseqn1980s:.\*  
2: Geneseqn1990s:.\*  
3: Geneseqn2000s:.\*  
4: Geneseqn2001as:.\*  
5: Geneseqn2001bs:.\*  
6: Geneseqn2002s:.\*  
7: Geneseqn2003as:.\*  
8: Geneseqn2003bs:.\*  
9: Geneseqn2003cs:.\*  
10: Geneseqn2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5134	100.0	2869	7 ABZ77632	Nucleotid
2	5134	100.0	2869	7 ADD22922	Acidother
3	2174	42.3	2823	7 ABV76941	Nucleotid
4	1558.5	30.4	2849	9 ADD42060	Trichoder
5	1167.5	22.7	3668	9 ADD24893	Trichoder
6	940	18.3	5698	9 ADD42054	Trichoder
7	823	16.0	3687	7 ABZ77634	Nucleotid
8	702	13.7	2289	7 ABZ77633	Nucleotid

9	429	8.4	3365	7 ABZ776162	Abz776162 A. cellul
10	409	8.0	11706	2 AAX55661	Aax55661 DNA seque
11	409	8.0	11707	6 AAD26525	Aad26525 Active ce
12	392.5	7.6	1103	3 AAF14988	Aaf14988 Trichoder
13	365.5	7.1	651	5 AAS84956	Aas84956 DNA encod
14	345	6.7	18596	4 AAF31109	Aaf31109 Thymidyla
15	345	6.7	18596	5 AAC91215	Aac91215 Human thy
16	345	6.7	18596	6 ABL67927	Abi67927 Ovary can
17	345	6.7	18596	6 ABL62854	Abi62854 Breat ca
18	345	6.7	18596	6 ABL63078	Abi63078 Breat ca
19	345	6.7	18596	6 ABK43334	Abk43334 Human Thy
20	345	6.7	18596	6 ABN95092	Abn95092 Gene #159
21	334.5	6.5	6415	2 AAX55662	Aax55662 DNA seque
22	334	6.5	6416	6 AAD26526	Aad26526 Active ce
23	330.5	6.4	7407	7 ACA42281	Aca42281 Prokaryot
24	330.5	6.4	7407	8 ACC59398	Acc59398 Microbial
25	323	6.3	2600	2 AAQ15178	Aaq15178 Portion o
26	318	6.2	1080	6 ABI99537	Abi99537 Mouse isc
27	302	5.9	1050	4 AAD21685	Aad21685 Mutationa
28	302	5.9	2849	4 AAD21684	Aad21684 Human ret
29	296.5	5.8	4767	3 AAA14667	Aaa14667 Nucleotid
30	293	5.7	4818	3 AAA14669	Aaa14669 Nucleotid
31	290	5.6	1232	4 AAD11124	Aad11124 Human sma
32	289	5.6	324	7 ACA25426	Aca25426 Prokaryot
33	288.5	5.6	4571	3 AAA14662	Aaa14662 Nucleotid
34	288	5.6	288	6 ABL51804	Abi51804 Hydroxypr
35	285.5	5.6	5184	8 ACC80175	Acc80175 Human MC2
36	282.5	5.5	9903	7 ACA40315	Aca40315 Prokaryot
37	282.5	5.5	110000	4 AA199682_04	Continuation (5 of
38	281.5	5.5	432	4 AA156352	Aai56352 Probe #25
39	281.5	5.5	518	4 AA143371	Aai43371 Probe #12
40	281.5	5.5	110000	4 AA199683_37	Continuation (38 o
41	281.5	5.5	110000	6 ABS55320_0	Abas55320 Human tra
42	281.5	5.5	341511	6 ABS55200	Abas55200 Genomic D
43	280.5	5.5	3008	7 AAD55730	Aad55730 Argiope t
44	280	5.5	49999	2 AAZ23895	Aaz23895 Murine LO
45	279	5.4	4547	3 AAA14664	Aaa14664 Nucleotid

ALIGNMENTS

RESULT 1

ABZ77632

ID ABZ77632 standard; DNA; 2869 BP.

XX ABZ77632;

AC ABZ77632;

DT 03-JUN-2003 (first entry)

DE Nucleotide sequence of the avicelase AvIII.

XX Avicelase; AvIII; glycoside hydrolase; enzyme; cellulase; biofuel;

XX detergent; pulp processing; paper processing; feed processing; textile;

XX cellulose; gene; ss.

OS Acidothermus cellulolyticus.

XX Key Location/Qualifiers

PH CDS 1..2869

FT /\*tag= a

FT /partial

FT /product= "AvIII"

FT /transl\_except= (pos:2869,aa:Xaa)

FT /note= "Xaa is an unspecified residue"

XX WO2003012090-A2.

XX 13-FEB-2003.

XX 28-JUL-2001; 2001WO-US023818.

XX 28-JUL-2001; 2001WO-US023818.

PA (WIDE ) MIDWEST RES INST.  
 XX Ding S, Adney WS, Vinzant TB, Himmel MB;  
 XX WPI; 2003-248177/24.  
 DR P-PSDB; ABP73015.  
 XX New thermostable Aviii peptide from Acidothermus cellulolyticus, useful  
 PT for degradation of cellulose or in generating anti-Aviii antibodies for  
 PT purifying recombinant Aviii polypeptides from genetically engineered  
 PT host cells.  
 XX  
 XX Claim 3; Page 24; 44pp; English.  
 XX  
 XX The present sequence encodes a thermostable avicelase polypeptide,  
 CC designated Aviii. Aviii is a member of the glycoside hydrolase family  
 CC of enzymes, and is a cellulase. Aviii is useful in the conversion of  
 CC biomass to biofuels and biofuel additives. It may be useful in the  
 CC production of detergents, pulp and paper processing, food and feed  
 CC processing and in textile processes. The thermostable Aviii peptide is  
 CC useful in the degradation of cellulose, and in generating specific anti-  
 CC Aviii antibodies that are useful in purifying recombinant Aviii  
 CC polypeptides from genetically engineered host cells, in detecting Aviii  
 CC polypeptide expression, as well as a reagent tool for characterizing the  
 CC molecular actions of the polypeptide. The Aviii polynucleotide is useful  
 CC as a source of probes or primers in various diagnostic assays  
 XX  
 XX Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;  
 SQ

Alignment Scores:  
 Pred. No.: 1,12e-172 Length: 2869  
 Score: 5134.00 Matches: 956  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.98% Indels: 0  
 DB: 7 Gaps: 0

US-09-917-376-1 (1-957) x ABZ77632 (1-2869)

QY 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgLeuValSerLeuLeu 20  
 DB 1 ATGGATCGTTCGGAGAACATCCGCTGCTGACTATGAGATCAGACGATTTGGTATCTACTGCTC 60  
 QY 21 AlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThr 40  
 DB 61 GCCGCCACTCGCTGCTCGCGTGGCGCGCTCTGGAGTCTTCCCATCGCGATAACG 120  
 QY 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGly 60  
 DB 121 GCTTCTCTCGGCACGCGCGGAGCTACGCGGTACACCTGGAGCAACGTTGGCGATCGGG 180  
 QY 61 GlyGlyGlyPheValAlaGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80  
 DB 181 GCGCGCGCGCTTTGTGACGGGATCGTCTTCAATGAAGGTGACCGGAAATCTGTAGCTG 240  
 QY 81 ArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu 100  
 DB 241 CGACGGACATCGGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
 QY 101 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp 120  
 DB 301 CTGGATTTGGTGGATGGAACAAATTTGGGTGTAACAACGCGCTGTCAGCATTTGGCGAGAC 360  
 QY 121 ProIleAsnThrAsnIleValTrpAlaValAlaGlyMetTyrThrAsnSerTrpAspPro 140  
 DB 361 CCATCAATCACTAACAGATGATGGCGCGCTGCGAATGTACACCAACAGCTGGGACCCA 420  
 QY 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu 160  
 DB 421 AACGACGGAGCATTTCTCCGCTGCTGTGATCAGGCGCAACGTTGGCAATAACGCCCTTG 480  
 QY 161 ProPheIleLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180

DB 481 CCGTTCAAGCTTGGCGGCAACATCCCGGGCGTGGATGGAGCGGCTTGGGTGGAT 540  
 QY 181 ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSer 200  
 DB 541 CCAAAACAATGACAAACATTTCTGTAATTTTCGGCGCCCGAGCGGCAAGGGCTCTGGAGAAGC 600  
 QY 201 ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle 220  
 DB 601 ACAGATTCGGCGGACCTGCTCCAGATGACGAACATTTCCGGAGCTAGGACGATACATT 660  
 QY 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla 240  
 DB 561 GCAATATCCCACTGACACGCGGCTATCAGACGATATTCAGGGGCTGCTTGGGTGCT 720  
 QY 241 PheAspIleSerSerSerSerLeuGlyGlnIleAspIleThrIlePheValGlyValAla 260  
 DB 721 TTCGACAAATCTTCGTCATCGCTCGGCAAGCGAGTAAGACCATTTTGTGGCGTGGCG 780  
 QY 261 AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnIleValPro 280  
 DB 781 GATCCCAATAATCCGGTCTTCTGGAGCAGACGCGCGCGGACGCTGGCGAGCGGTGGCG 840  
 QY 281 GlyAlaProThrGlyPheIleProHisIleValPheAspProValAsnHisValLeu 300  
 DB 841 GGTGCGCGGACCGGCTTCCCGCAGAGGCGCTTTGACCCCGGTCAACACGCTGCTC 900  
 QY 301 TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpIle 320  
 DB 901 TATATGCGCACCAATACGCGTGGTCCGTATGACGGGAGCTCCCGCGACGCTCGGAAA 960  
 QY 321 PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla 340  
 DB 961 TTTCTCGGTGACCTCCGGGACATGAGCGGAATCAGCCGCTACCTTCGACGACACGCGCC 1020  
 QY 341 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360  
 DB 1021 AACGACTACTTTGGTTACAGCGGCTCTACTATCGACCGCGGACCCGCAACAGATAATG 1080  
 QY 361 ValAlaThrGlnIleSerTrpProAspThrIleIlePheArgSerThrAspGlyGly 380  
 DB 1081 GTGGCAACCCAGATATCTGTGTGGCGGACACCAATAATCTTTCGGAGCACCGCGCGGT 1140  
 QY 381 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 400  
 DB 1141 CGCAGCTGGACGCGGATCTGGGATTCGAGCTTATCCCAATCGAAGCTTGGATATGTG 1200  
 QY 401 LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro 420  
 DB 1201 CTTGACATTTCCGGGAGGCTTGGCTGACCTTCGGCGGTACAGCCGAATCTCCCGTACCC 1260  
 QY 421 SerProIleLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 440  
 DB 1261 AGTCGAGCTCGGCTGGATGGATGAGCGATGGCAATCGATCCGTTCAACTCTGATCGG 1320  
 QY 441 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrIleTrpAspSer 460  
 DB 1321 ATGCTCTACGGAACAGCGCGGACGCTGTAGCGCAACAATGATCTCAGAAATGGGACTCC 1380  
 QY 461 GlyGlyGlnIleHisIleAlaProMetValIleGlyLeuGluGluThrAlaValAsnAsp 480  
 DB 1381 GCGCGCCAGATTCATATCGCGCGATGGTCAAAAGATTGGAGGAGACGCGCGTAAACGAT 1440  
 QY 481 LeuIleSerProSerGlyValAlaProLeuIleSerAlaLeuGlyAspLeuGlyIlePhe 500  
 DB 1441 CTCATCAGCCCGGCTTGGCGGCCCCGCTCATCAGCGCTCTCGAGACCTCGCGGCTTC 1500  
 QY 501 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 520  
 DB 1501 ACCCAGCGCGAGTTCCTGCGGTGCTCAGCATCTTCCAGCTCACCCTGCTTTCAGACC 1560  
 QY 521 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 540  
 DB 1561 GGCACCGCGTTCGACTATCGGAATTTGAATCCGTCGATCATCGTTTCGCGCTGGAAGTTTC 1620



```
QY 541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 560
DB 1621 GATTCATCGAGCCAAACCGAACGACAGCGTCGCGTTCTTCGACGACGCGCGCAAGAAC 1680
QY 561 TrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaLysAla 580
DB 1681 TGGTTCCAAAGGACGACGACCTGGCGGGGTGACACGGGGCGGACCGTCGCGCATCGGCC 1740
QY 581 AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTrpAlaVal 600
DB 1741 GACGGCTCTCGTTTCTGCTGGGCTCCCGCGCATCCCGGTGACGCTGTGTGTACGAGTC 1800
QY 601 GlyPheGlyAsnSerTrpAlaLysGlnGlyValProAlaAsnAlaGlnIleArgSer 620
DB 1801 GGATTTGGCAACTCTCTGGGCTGTTCCGCAAGGTGTTCCCGCCAAATCCCGCAGATCCGCTCA 1860
QY 621 AspArgValAsnProLysThrPheThrAlaLeuSerAsnGlyThrPheThrArgSerThr 640
DB 1861 GACCGGGTGAATCCAAAGACTTTCTATGCCCTATCCAAATGGAACTTCTATCGAAGCAG 1920
QY 641 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 660
DB 1921 GACCGCGCGGTGACATTCACACGGTCCGCGCGGCTTCCGAGCAGCGTCCGCTCGGT 1980
QY 661 ValMetPheHisAlaValProGlyLysGlyAspLeuTrpLeuAlaLysSerGly 680
DB 1981 GTCATGTTCACGCGGTCTCGAAGAGGAGCGATCTGTGGCTCGCTCATCGAGCGGG 2040
QY 681 LeuThrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAla 700
DB 2041 CTTTACCACTCAACCAATGGCGGACAGTGTGCTGCAATCACCGCGGTATCTCTCCGG 2100
QY 701 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTrpProAlaValPheValVal 720
DB 2101 GTGAACGTGGGATTTGGTAAGTCTCGCGCCGGGTGTCATACCACCGCTCTTTGTGTC 2160
QY 721 GlyThrIleGlyValThrGlyAlaThrArgSerAspAspCysGlyThrTrpVal 740
DB 2161 GGCAAGATCGAGGCGTTACGGGGGGGTACCGCTCCGACGACGTGGGACGACCTGGGTA 2220
QY 741 LeuIleAsnAspAspGlnHisGlnTrpGlyAsnTrpGlyAlaIleThrGlyAspHis 760
DB 2221 CTGATCAATGATGACGACCAATACGGAATATGGGGACAGCAATACACCGGTGACCAC 2280
QY 761 AlaAsnLeuArgArgValTrpIleGlyThrAsnGlyArgGlyIleValTrpGlyAspIle 780
DB 2281 GCGAATTTACGCGGGGTGTACATAGGCACGAACGCGCGTGGAAATGTGTATACGGGGACATT 2340
QY 781 GlyGlyAlaProSerGlySerProSerProSerValSerProSerSerAlaSerProSerLeu 800
DB 2341 GGTGTGTGCGCTCGGATCGCGCTCTCGCTCGGTGAGTCCGTCGGCTTCGCGAGCCCTG 2400
QY 801 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820
DB 2401 AGCCCGAGCCCGAGCCGAGCAGCTCGCCATCGCGCTCGCGCTCGCGAGCTCGAGTCCA 2460
QY 821 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840
DB 2461 TCTCTGTCGCGCTCTCTCGTCGCGCTCACCATCGCGGAGTCCGTCGCTCGCTCCGTCACCA 2520
QY 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860
DB 2521 TCGGCTGTCGCGAGCCGCTCTCTGTCACGAGCCGCTCTCTGTCACCGCTCTCTGTCGCGCG 2580
QY 861 SerProThrProSerSerProValSerGlyValLysValGlnTrpLysAsnAsn 880
DB 2581 AGCCCAACGCGCTCTGTCGCGGTGTGGGTGGGGTGAAGGTGAGTATAGATATAT 2640
QY 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900
DB 2641 GATTTGCGCGCGGGTGATATATCAGATCAAGCGCGGTTTTCAGGTGTGTGAATACCGGGTCTG
```

```
QY 901 SerSerValAspLeuSerThrValThrValArgTrpPheThrArgAspGlyGlySer 920
DB 2701 TCGTCGGTGATTTGTGACGGTGACGGTGCGGTACTGGTTTACCCGGGATGGTGCTCG 2760
QY 921 SerThrLeuValTrpAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSer 940
DB 2761 TCGACACTGGTGACAACTGTGACTGGCGGCGGATCGGGTGTGGGAATATCCGCGCTCG 2820
QY 941 PheGlySerValAsnProAlaThrProThrAlaAspThrTrpLeuGln 956
DB 2821 TTGCGCTCGGTGAACCGCGGACGCCGACGCGGACACCTACTACCTGCAG 2868

RESULT 2
ADD22922
ID ADD22922 standard; DNA; 2869 BP.
XX
AC ADD22922;
XX
DT 15-JAN-2004 (first entry)
XX
DE Acidothermus cellulolyticus avicelase AvIII DNA.
XX
KW AvIII; cellulose reduction; agricultural biomass; municipal solid waste;
KW glycoside hydrolase; avicelase; ds; gene.
XX
OS Acidothermus cellulolyticus.
XX
FH Key Location/Qualifiers
FT CDS 1..2868
FT /*tag= a
FT /product= "AvIII"
XX
PN US2003108988-A1.
XX
PD 12-JUN-2003.
XX
PF 18-OCT-2002; 2002US-00155400.
XX
PR 28-JUL-2001; 2001US-00917376.
XX
PA (DING/) DING S.
PA (ADNE/) ADNEY W S.
PA (VINZ/) VINZANT T B.
PA (HIMM/) HIMMEL M E.
XX
PI Ding S, Adney WS, Vinzant TB, Himmel ME;
XX
DR WPI; 2003-810853/76.
DR P-PSDB; ADD22921.
XX
PT New isolated thermal tolerant avicelase polynucleotide useful for
PT detection of a polynucleotide encoding AvIII and for reducing cellulose
PT in a starting material, e.g. municipal solid waste.
XX
XX Claim 17; SEQ ID NO 2; 29pp; English.
XX
CC The invention relates to an isolated polynucleotide molecule encoding a
CC thermostable AvIII polypeptide. The polynucleotide is useful for
CC detection of a polynucleotide encoding AvIII. The polynucleotide is
CC useful for reducing cellulose in a starting material which involves
CC administering to the starting material, e.g. agricultural biomass or
CC municipal solid waste, a polypeptide molecule of the polynucleotide. The
CC method further comprises administering a second polypeptide molecule
CC chosen from the glycoside hydrolase family of proteins. The present
CC sequence represents DNA encoding the Acidothermus cellulolyticus
CC avicelase AvIII.
XX
SQ Sequence 2869 BP; 546 A; 857 C; 900 G; 565 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.: 1.12e-172 Length: 2869
Score: 5134.00 Matches: 956
Percent Similarity: 100.00% Conservative: 0
```



Best Local Similarity: 100.00%		Mismatches: 0
Query Match: 99.98%		Indels: 0
DB: 9		Gaps: 0
US-09-917-376-1 (1-957) x ADD22922 (1-2869)		
Qy	1 MetAspArgSerGluAanlleAArgLeuThrMetArgSerArgArgLeuValSerLeuLeu	20
Db	1 ATGGATCGTTGGAGAAACATCCGCTCAGCTATGAGATCAGACGATTGGTATCACTGCTC	60
Qy	21 AlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThr	40
Db	61 GCGCCCACTCGCTGCTTGGCGTGGCGCGCTCTGGAGTTCTGCCCATCGCGATACG	120
Qy	41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTpsSerAsnValAlaIleGly	60
Db	121 GCTTCTCTCGGCACGCGCGGACACTCAGCGGTACACCTGGAGCAAGTGGCGATCGG	180
Qy	61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal	80
Db	181 GCGCGCGGCTTTGTTCGACGGGATCGTCTTCAATGAAGGTGCACCGGGAATTTCTGTACGTG	240
Qy	81 ArgThrAspIleGlyMetTyrArgTpsAspAlaAlaAsnGlyArgTpsIleProLeu	100
Db	241 CGGACGGACATCGGGGGATGTATCGATGGATGCGGCCAACCGCGGTGGATCCCTCTT	300
Qy	101 LeuAspTpsValGlyTpsAsnAsnTpsGlyTyrAsnGlyValValSerIleAlaAlaAsp	120
Db	301 CTGGATTGGTGGATGGACAAATTTGGGGTACACCGCGCTGTCAGCATTTGGCGAGAC	360
Qy	121 ProIleAsnThrAsnLysValTpsAlaAlaValGlyMetTyrThrAsnSerTpsAspPro	140
Db	361 CCGATCAATACTAACAAAGTATGGCGCGCTCGGAATGTACACCAACAGCTGGGACCA	420
Qy	141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTpsGlnIleThrProLeu	160
Db	421 AACGACGGACGATCTCCGCTCGTCTGTATCAGGGCGCAACGTGGCAATAAAGCCCTG	480
Qy	161 ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp	180
Db	481 CCGTTCAAGCTTGGCGCAACATGCCCGGGGTGGAATGGCGAGCGGCTTGGGTGGAT	540
Qy	181 ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTpsArgSer	200
Db	541 CCAAAACAATGACAAACATTTCTGATTTTCGGCGCCCGAGCGCAAGGGCTCTGGAGAAGC	600
Qy	201 ThrAspSerGlyAlaThrTpsSerGlnMetThrAsnPheProAspValGlyThrTyrIle	220
Db	601 ACAGATTCCGGCGGACCTGCTCCAGATGACCAACTTTCCGGACGTAGGCACGTACATT	660
Qy	221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTpsValAla	240
Db	661 GCAATCCCACTGACACGACCGGCTATCAGAGCGATATTCAAGGCGTGTCTGGGTGCT	720
Qy	241 PheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAla	260
Db	721 TTCGACAAGTCTTCGTATCGCTCGGCAAGCGAGTAAGACCAATTTTGTGGGCGTGGCG	780
Qy	261 AspProAsnAsnProValPheTpsSerArgAspGlyGlyAlaThrTpsGlnAlaValPro	280
Db	781 GATCCCAATAATCCGGTCTTCGAGCAGAGACGCGCGCGCGACGTGGCAGCGCGTCCG	840
Qy	281 GlyAlaProThrGlyPheIleProHisGlyValPheAspProValAsnHisValLeu	300
Db	841 GGTGCGCGGACGGCTTCATCCCGCACAAGGGCGTCTTTGACCGCGTCAACACGCTGCTC	900
Qy	301 TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTpsLys	320
Db	901 TATATTGCCACCAACATACGGTGGTTCGTATGACGGGAGCTCCGCGAGCTCGAAA	960
Qy	321 PheSerValThrSerGlyThrTpsThrArgIleSerProValProSerThrAspThrAla	340
Db	961 TTCTCGGTGACCTCCGGGACATGACGCGGAATCAGCCCGGTACCTTCGACGGACACGGCC	1020
Qy	341 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet	360
Db	1021 AACGACTACTTTGGTTACAGCGGCTCACTATCGACCGCCAGCACCCGAAACGATAATG	1080
Qy	361 ValAlaThrGlnIleSerTpsTpsTpsProAspThrIleIlePheArgSerThrAspGlyGly	380
Db	1081 GTGGCAACCCAGATATCGTGTGGCGGACACCAATAATCTTTTCGGAGCACCCGACGCGGT	1140
Qy	381 AlaThrTpsThrArgIleTpsAspTpsThrSerTyrProAsnArgSerLeuArgTyrVal	400
Db	1141 GCGAGCTGGACCGGATCTGGGATTCGAGGATATCCCAATCGAAGCTTCGATATG	1200
Qy	401 LeuAspIleSerAlaGluProTpsTpsPheGlyValGlnProAsnProProValPro	420
Db	1201 CTTGACATTTCCGGGAGCCTTGGCTGACCTTCGGCTACAGCCGAATCTCCCGTACCC	1260
Qy	421 SerProLysLeuGlyTpsMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg	440
Db	1261 AGTCCGAAGCTCGGCTGGATGGATGAAGCATGGCAATCGATCCGTTCAACTCTGATCGG	1320
Qy	441 MetLeuTyrGlyThrGlyValaThrLeuTyrAlaThrAsnAspLeuThrLysTpsAspSer	460
Db	1321 ATGCTCTACGGAACAGGCGCGAGCTTGTACGCAACAATGATCTCAGAAGTGGGACTCC	1380
Qy	461 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValaAsnAsp	480
Db	1381 GCGCGCCAGATTATCATTCGCGCGATGTTCAAGATTGGAGAGACGCGCGTAAACGAT	1440
Qy	481 LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe	500
Db	1441 CTCATCAGCCCGCGCTCTGGCGCCCGCTCATCAGCGCTCTCGGAGACCTCGCGCGCTTC	1500
Qy	501 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr	520
Db	1501 ACCACCGCGAGCTTACTGCGCTCCATCGAGATCTTCACGTACCGGTGTTACGACC	1560
Qy	521 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe	540
Db	1561 GGCACCGCTCGACTATCGGAATTTGAATCGTCTGATCATCGTTCGCGCTGGAAGTTTC	1620
Qy	541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn	560
Db	1621 GATCCATCGAGCAACCCGACGACGAGCGACGCTCGCGTTCTCGACAGACGCGCGCGAAGAC	1680
Qy	561 TrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAlaSerAla	580
Db	1681 TGGTTCCAAAGCAGCGAACCTTGGCGGGGTGACGACGGGGCGGACCGTCCCGCATCGGCC	1740
Qy	581 AspGlySerArgPheValTpsAlaProGlyAspProGlyGlnProValValTyrAlaVal	600
Db	1741 GACGCTCTCGTTTCGTTCTGGGCTCCCGCGCATCCGGTACGCTGTGGTGTACGACGTC	1800
Qy	601 GlyPheGlyAsnSerTpsAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer	620
Db	1801 GGATTTGGCAACTCTCGGCTGCTTCGCAAGGTTCGCGCAATGCCAGATCCGCTCA	1860
Qy	621 AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr	640
Db	1861 GACCGGGTGAATCCAAAGACTTTCTATGCCCTATCCCAATGGAACTTCTTATCGAAGCAG	1920
Qy	641 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly	660
Db	1921 GACGGCGGCTGACATTCCAACCGGTCGCGCGGCTCTTCGAGGAGCGGTCGCTCGGT	1980
Qy	661 ValMetPheHisAlaValProGlyLysGlyAspLeuTpsLeuAlaAlaSerSerGly	680
Db	1981 GTCATGTTCCACGCGGCTGCGCTGGAAAGAGCGATCTGTGGCTCGTGCATCGAGCGG	2040
Qy	681 LeuTyrHisSerThrAsnGlyGlySerSerTpsSerAlaIleThrGlyValSerSerAla	700
Db	2041 CTTTACCACCTACCAATGGCGGACGAGCTTGGTCTGCAATCAGCGGGGTATCTCTCCGG	2100

```

QY 701 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 720
Db 2101 GTGAACGTGGGATTTGGTAAGTCTGCCCGCGGGTCGTATACCCAGCGCTTTGTCTGC 2160
QY 721 GlyThrIleGlyGlyValThrGlyValTyrArgSerAspAspCysGlyThrThrVal 740
Db 2161 GGCACGATCGAGCGGTTCAGGGGCGGTACCGCTCCGACGACTGTGGACGACTGGGTA 2220
QY 741 LeuIleAsnAspAspGlnHisGlnTyrGlyAsnTyrGlyGlnAlaIleThrGlyAspHis 760
Db 2221 CTGATCAATGATGACACGACCAATACGAAATTGGGACAAAGCAATCACCGGTTGACCAC 2280
QY 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
Db 2281 GCGAATTTACGGCGGGTGTACATAGGCACCAAGACCGCGGTGGAATTGTATACGGGGACATT 2340
QY 781 GlyGlyAlaProSerGlySerProSerProSerProSerProSerProSerProSerProSer 800
Db 2341 GGTGGTGGCGCGTCCGATCGCGTCTCCGTCGGTGAAGTCGGTCCGCTTCGCCAGCGCTG 2400
QY 801 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 820
Db 2401 AGCCGAGCGCGACCGCGAGCAGCTCCCATCGCCGTCGCCGTCGCCGAGCTCGAGTCCA 2460
QY 821 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 840
Db 2461 TCCGTCGCGCGTCTCCGTCGCGTCCATCGCGAGTCCGTCGCTCGCTCCGTCACCA 2520
QY 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 860
Db 2521 TCGCGCTCGCGAGCGGTCTTCGTACCGAGCGCGTCTTCGTCCACCGTCTTCGTCCGCG 2580
QY 861 SerProThrProSerProSerProSerProSerProSerProSerProSerProSerProSer 880
Db 2581 AGCCCAACGCGCTCGTCCGCGTCTCGCGTGGGTGAAGTGCAGTATAAGAAATAAT 2640
QY 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900
Db 2641 GATTTCGGCGCGGTGATATATCATGATCAAGCGGCTTTCAGGTGGTGAATACCGGCTCG 2700
QY 901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 920
Db 2701 TCGTCGGTGGATTTGTCCAGCGTGACGCTGCGGTACTGGTTTCAACCGGAGTGGTGGCTG 2760
QY 921 SerThrLeuValTyrAsnCysAspTrpAlaIleGlyCysGlyAsnIleArgAlaSer 940
Db 2761 TCGACATGTTGTACAACCTGTGACTGGCGCGCGATCGGGTGGGAATATCCGCGCTCG 2820
QY 941 PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 956
Db 2821 TTCGGCTCGGTGAACCGCGGACGCCGACCGCGGACACCTACCTGCGAG 2868

RESULT 3
ABV76941
ID ABV76941 standard; DNA; 2823 BP.
XX AC ABV76941;
XX DT 03-MAR-2003 (first entry)
XX DE Nucleotide sequence of a xyloglucanase enzyme.
XX KW Xyloglucanase; enzyme; family 74; glycosyl hydrolase; cellulosic fiber;
XX OS Jonesia sp.
XX PH Location/Qualifiers
FT CDS 1..2823
FT /product= a
FT /tag= "xyloglucanase enzyme"
FT sig_peptide 1..69
FT /tag= b

```

```

mat_peptide 70..2820
/*tag= c
WO200277242-A2.
03-OCT-2002.
27-MAR-2002; 2002MO-DK000210.
27-MAR-2001; 2001DK-00000504.
(NOVO ) NOVOZYMES AS.
Duffner F, Sjolholm C;
P-PSDB; ABB99489.
WPI; 2003-092855/08.
New xyloglucanase enzyme belonging to family 74 of glycosyl hydrolases,
and endogenous to a bacterium, useful in the textile industry for
improving properties of cellulosic fibers, yarn or fabric.
Claim 11; Page 68-72; 76pp; English.
The present sequence encodes a xyloglucanase enzyme, belonging to family
74 of glycosyl hydrolases. The enzyme is isolated from Jonesia sp.
DSM14140. The enzyme is useful in processes for machine treatment of
fabrics. It is also useful in the textile industry for improving the
properties of cellulosic fibers, yarn, woven or non-woven fabric, and in
a textile scouring process step. The xyloglucanase enzyme is also useful
in the cellulose fiber processing industry for rating of fibers such as
hemp, jute, flax and linen. It is useful for preventing binding of
certain soils to the xyloglucan left on the cellulosic material
XX SQ Sequence 2823 BP; 665 A; 803 C; 764 G; 591 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,99e-68 Length: 2823
Score: 2174.00 Matches: 430
Percent Similarity: 59.48% Conservative: 141
Best Local Similarity: 44.79% Mismatches: 311
Query Match: 42.34% Indels: 78
DB: 7 Gaps: 16

US-09-917-376-1 (1-957) x ABV76941 (1-2823)
QY 10 ThrMetArgSerArgArgLeuValSerLeuLeuAlaIleThrAlaSerPheAlaValAla 29
Db 13 ACCGTGGTGGCCACAGCACTAACTCGCTGCTCGCCACGACTGCCGTGGCATCGCGCC 72
QY 30 AlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaIleAlaIleThrThr 49
Db 73 ACATCA-----ACCCCATCGCTAGTAACTGACCGCCCTCTTCTTCATCATCACT 123
QY 50 GlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyGlyPheValAspGlyIleVal 69
Db 124 TCCGGCTATAGCTGTGACAGTTGAGATTGTGGCGCGCGGTTTTGTGCTCGAATTGTC 183
QY 70 PheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyIleGlyMetTyrArg 89
Db 184 TTTAACCCAGAAAGACCCAGGCTTGGTGTATGCCGACAGATATTGGTGGACGCGTACCGG 243
QY 90 TrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrp 109
Db 244 CTCAATGACTCCACCGCGGTGGATCCCTCTTACGGACCATATTGGTTGGGATGACTGG 303
QY 110 GlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIleValTrpAla 129
Db 304 AGCCACTCCGGAATTCTGTCTCTGCGCACTGACCTGTGTGACACCAACCGCGTGTACCTT 363
QY 130 AlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSer 149
Db 364 GCAGCGGACGTAATCCATGGACTGGATCCACAAACGGTGTCTATTGTTGGTTCGCGG 423

```

Qy	150	AspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLeuGlyGlyAsnMetPro	169
Db	424	GACAAAGCGCAGACCTGGGAAAGACGATGCTGCGGTTCGGTGGGAGGCAACATGCCG	483
Qy	170	GlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPhe	189
Db	484	GGTGGCGGGATGGGAGAACGCCCTCGCGGTGTGACCCAAACAATAAAGTCTCTGTACTTT	543
Qy	190	GlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGln	209
Db	544	GGTGGGAATCGGGTAATGACATATGGAAGTCAACTGACTACGGGAAAACTGGGGCAA	603
Qy	210	MetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGly---	228
Db	604	GTGACGAGCTTCCCAATGCTGTGTAATCTAGCTGCC-----GATGCTTCTGTGTGCG	654
Qy	229	TyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeu	248
Db	655	TACACGGGCAAAACCAAGGTGTGTGGGTAACTTTGACCCCACTCAGCGAAAGCT	714
Qy	249	GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrp	268
Db	715	GGTCAGACAAACAAACGATCTAGTAGGTGTAGCAGCAACAGCAACAACGTTTATCGA	774
Qy	269	SerArgAspGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIlePro	288
Db	775	TCTACCGATGGCGGGCAACATGCGAACCGCTCCCTGGCAACCCACCGGATTCTCTGGCG	834
Qy	289	HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly	308
Db	835	CAAAAGGGAGTGTGACCAATAAAGTCAACCACTCTACATCGCTACTCAGATACGTGGT	894
Qy	309	GlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrp	328
Db	895	GGCCCGTATCAGCGGTCAAGGGGGATGTGTGGCGCTAGATATTTCCAGTGGCCAGTGG	954
Qy	329	ThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGly	348
Db	955	ACTCGAATCAGCCCAATTCCTCTAC- --TCATCGAATTACGGGTTCGGGTACTCCGGT	1011
Qy	349	LeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrp	368
Db	1012	TTAGCCATTGACCGCAAAACCTTGACACCATCATGTTGTTTCCCAAGTCTTGTGTGG	1071
Qy	369	ProAspThrIleIlePheArgSerThrAspGlyAlaThrTrpThrArgIleTrpAsp	388
Db	1072	CCTGCATGTACGTATTCGTCTCCACGACGGGGGAAAAAGTGTGCGCTATCTGGGAA	1131
Qy	389	TrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrp	408
Db	1132	CTGAAACGGTTCACAGCGCGGTACAAACGAGTACACCAAGTACTCAGGGGTCCCTGG	1191
Qy	409	LeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrpMetAsp	428
Db	1192	CTTGACTTTGGCAATACAGCCAAAGACCTGAAGCTAAACCCCAAACTCGGTGGATGACC	1251
Qy	429	GluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThr	448
Db	1252	CAGTCTTCGAGATTGACCGCGATAACTCCGACCGCTTCTTCTATGGCACCGGCGCGGC	1311
Qy	449	LeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaPro	468
Db	1312	ATCTATGGCGGAACCACTCAAACTGGATNAGGGGCAAAAANGTTGATATCACTGTC	1371
Qy	469	MetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSerGlyAla	488
Db	1372	AAGGCCCAAGGATCGAAGAAACCGCAGCCCAAGACCTTCGCGCTCCCTCGGAACATT	1431
Qy	489	ProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal	508
Db	1432	GGCCTATATCATCGGTGGGAGACATCGGTGGTTCACCCACAAGATATCTCTCAGGTG	1491

Qy	509	ProSerThr---llePheThrSerProValPheThrThrGlyThrSerValAspTyrAla	527
Db	1492	CCTAACAAAGTACTACTACAAGAAGACCCACACCATGACACCGGTGACCTCCATTAAGTACTGC	1551
Qy	528	GlulLeuAsnProSerIleValArgAlaGlySerPheAspProSerSerGlnProAsn	547
Db	1552	GAAGCAACCGCACTGCTGGTGGTGCAGGCAAA-----TCTATTTCTGGGGAACCC	1605
Qy	548	AspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluPro	567
Db	1606	ACCTCATGGGTTGGGGTTTCAACTGATGCTGGCGCAACAGTGGAAACCTGGCGCCACGCCA	1665
Qy	568	GlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrp	587
Db	1666	TCAGGGGTTAAAGGCCCGCGCTCCATCAGTGTGTGACGTAATGCTTCATCCATTTGTGTGG	1725
Qy	588	AlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAla	607
Db	1726	GCCTCCG-----GAAGGTGGCGCTCTCTGACGTTTCCACCAAGTGGTCTCTCTGTGTCA	1779
Qy	608	AlaSerGlnGlyValProAlaAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThr	627
Db	1780	ACGGTGAGCGGGCTTCTCTCAATGCCAGGTCGCATCAGCCGAGTCAAGCTAACTACT	1839
Qy	628	PheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGln	647
Db	1840	CTTTACGGCTTCGTTGACGGAAAGTCTATCACTCCACAAATGTTGGGGCTTCCTTCACG	1899
Qy	648	ProValAla---AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal	666
Db	1900	CGGTACGCTTTCACGGGGTTTCCACACAGTGA-----AACGTTCTGTTCCGTCGCGTG	1953
Qy	667	ProGlyLysGluGlyAspLeuTrpLeuAla-----AlaSerSerGlyLeu	681
Db	1954	CCGGCGCTCAAGGTCACTCTGGCTGTGTTGGTGGATATCGGGAACGACCTATGGGATG	2013
Qy	682	TyrHisSerThrAsnGlyLysSerTrpSerAlaIleThrGlyValSerSerAlaVal	701
Db	2014	TGGCGGTCCACTGATGTCGGGAAGAACTGACCAAGGTACTGAGATTTCAGGAAGGCGAT	2073
Qy	702	AsnValGlyPheGlyLysSerAlaProGlyLysSerTyrProAlaValPheValValGly	721
Db	2074	CGCGTTCGGTTTCGGTAAAGCAACGAGTGGGATACCCAGTGATTTTCACATCCGCC	2133
Qy	722	ThrIleGlyValThrGlyValTyrArgSerAspAspCysGlyThrTrpValLeu	741
Db	2134	AAGATTATGGAGTTCAGGAATTTCCGCTCTGACGATGAAGCAAGACCTGGAAACGC	2193
Qy	742	IleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAla	761
Db	2194	ATCAACGACACCAACCAACCATGGTATTGACATGGCGCATCAATCACTCCGCTGACCCAGAT	2253
Qy	762	AsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGly	781
Db	2254	GTCGAAGGCGCGCTACTATGTTGATACCAACGGTCGGCGCATATCGTTGGTGA-----	2307
Qy	782	GlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSer	801
Db	2307	-----	2307
Qy	802	ProSerProSerProSerSerProSerProSerProSerProSerSerProSer	821
Db	2308	-----TCCTCCACTCCCGCCCGCGGTCAAGATAGCGGACGGCGATCG	2352
Qy	822	SerSerProSerProSerPro-----SerProSerProSerArgSerProSerPro	840
Db	2353	GGAAATACAGGGGCCACACCGAGCGCTGAAACCCCTGACAGCGATCAACACTGTGGCGGG	2412
Qy	841	SerAlaSerProSerProSerProSerProSerProSerSerProSerSerSerPro	860
Db	2413	TCCACTGATACGGGCGAAACTTCAGACCCGGGAACAGGAAGTGGATCTGGAACCGCACT	2472
Qy	861	SerProThrProSerSerProValSerGlyGlyValLysValGlnTyrLysAsnAsn	880



Qy	160	LeuProPheIysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal	179
Db	505	TTGGCCCTTCAAAGCTCGGGGGTAAATGCTCCAGGACGGAGCCGAGAGCGTCTGGCTGTC	564
Qy	180	AspProAsnAsnAspAsnIleLeuTyrPheGlyValaProSerSerGlyGlyLeuTyrArg	199
Db	565	GATCCGGCCCAACTCCAAACATCATCTACTTTGGTGGCTCGCTCAGGAAACGGCCCTCTGGAAG	624
Qy	200	SerThrAspSerGlyAlaThrTyrPheThrGlnMetThrAsnPheProAspValGlyThrTyr	219
Db	625	TCTACGACGGCGGGCTGACCTTTTCAGAGTCTCTGCTGTTTACGGCAACTGGGACGTAC	684
Qy	220	IleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTyrVal	239
Db	685	ATCCAGACCCGAGTGATTCACACGGCTCAACACAGCGACAAAGCAAGACTCATGTGGGTT	744
Qy	240	AlaPheAspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyVal	259
Db	745	ACGTTTCAGCTCAACACAGCAGCAGACCGGGGGAGCCACGTCTGTATCTTTTGTGGCAGC	804
Qy	260	AlaAspPro--AsnAsnProValPheThrSerArgAspGlyGlyAlaThrTyrGlnAla	278
Db	805	GCTGATAAACATCACTGCTTCAGTCTATGTAGCAGCAAGTCCCGCTCCACGTGAGTGCT	864
Qy	279	ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis	298
Db	865	GTACCGGGGAGCCAGGAAATACTTTCCTCACAGCGCAATGCGAGCCAGCAGAGAAG	924
Qy	299	ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspVal	318
Db	925	GCCTTGATCTGACCTATCCGATGGCACAGGCGCGTATGATGACACATGGCTCAGTG	984
Qy	319	TyrLysPheSerValThrSerGlyThrTyrThrArgIleSerProValProSerThrAsp	338
Db	985	TGGAGGTACGACATTCGAGGGGAACCTTGAAAGACATCACCCCTGCTCTGGATCATAG	1044
Qy	339	ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr	358
Db	1045	CTA-----TACTTTGGCTTGGCGGCTGGCCCTCGATTTGCAAAAGCCAGGACAC	1095
Qy	359	IleMetValAlaThrGlnIleSerTyrTyrProAspThrIleIlePheArgSerThrAsp	378
Db	1096	CTTGTTGTTGCTTCTTTTGAACCTCTTGTTGGCCAGATGCTCAGCTGTTTTCGTGTCACCGAC	1155
Qy	379	GlyGlyAlaThrTyrThrArgIleTyrAspTyrThrSerTyrProAsnArgSerLeuArg	398
Db	1156	TCTGGGCAACATGGAGCCCGATCTGGGCGTGGCGGAGCTATCCGACTGAGACC-----	1209
Qy	399	TyrValLeuAspIleSerAlaGluProTyrThrPheGlyValGln-----	414
Db	1210	TATTACTACAGCATCTCAGT-GATGCTACTTTAACGATCCGATCGGAATGGCACTGACTC	1268
Qy	415	-----ProAsnProProValPro-----	420
Db	1269	TGCTTCAAGACTCCCAAAGCACCGTGGATCAAGAACAACTTATTATCGATGTGACGAGCGAG	1328
Qy	421	SerPro-----LysLeuGlyTyrMetAspGluAlaMetAlaIleAsp	434
Db	1329	TCACCGTCCGATGGTCTCATCAGCGCTCGGCTGGATGATTTAGTCTCTCGAGATTGAC	1388
Qy	435	ProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAsp	454
Db	1389	CCAAACCAGACCAACCACTGGCTCTACGGCACCGGAATGAACTCTTTGGCGGCCACGAT	1448
Qy	455	LeuThrLysTyrAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGlu	474
Db	1449	CTACCAACTGGGACACGGCCCAATGTGTCAATCAATCACTGGCAGACGGCATCGAG	1508
Qy	475	GluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeu	494
Db	1509	GAATTCCTCCGAGGACCTGGCTCTGCAACCGCGGGAAGCGAGCTATTGGCGGCAGCT	1568

Qy	495	GlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValProSerThr	511
Db	1569	GSAGACACAAACGGCTTCACCTTTGGCCGACAGAAACGAACTCGGACATCGCCGACAGACG	1628
Qy	512	IlePheThrSerProValPheThrThrGlyThrSerValAspTyValaGluLeuAsnPro	531
Db	1629	GTCTGGGCAACGCCACATGGCCACCTCGACAGAGGTGCACTACGCGCGGAACCTCGGTC	1688
Qy	532	SerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisVal	551
Db	1689	AAGAGCGTCGTCGGCGTCGCGCAACACCGCGCGCACCAACAG-----GTG	1733
Qy	552	AlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThr	571
Db	1734	GCCATCTCGTCGACGCGCGCGAGCTGGAGATCGACTACCGCGCGCACAGCTCCCATG	1793
Qy	572	ThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAsp	591
Db	1794	AACGGCGGACAGGTGGCTATTGGCCGACGGGCACACGATCTCTTGCTCGACCGGCTCG	1853
Qy	592	ProGlyGlnProValValTyTrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGly	611
Db	1854	TCGGGC-----GTGCAGCGCTCGACGTTCCAGGGCAGCTTGCTCTCGTCTCGAGC	1904
Qy	612	ValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyTrAlaLeu	631
Db	1905	CTGCCCGCGGCGCGCTCATCGCTCGGACCAAGAACAACACGCTTCTTACGCGCGC	1964
Qy	632	SerAsnGlyThrPheTyTrArgSerThrAspGlyGlyValThrPheGlnProValAlaAla	651
Db	1965	TCGGGATCGACCTTTTACGTCTCAGCAAGGACACCGGCAGCGCTTC-----ACGCGC	2015
Qy	652	GlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValProGlyLys	669
Db	2016	GGG---CCCAAGCTGGCGAGCGGAGGACGATCGGGGATATCGCTGCTCACCCGACCA	2072
Qy	670	GluGlyAspLeuTrpLeuAlaAspSerGlyLeuTyHisSerThrAsnGlyGlySer	689
Db	2073	GCGGGCACGTTGTATGTCTCGACCGACGCTCGGCATATTCGCTCCACAGACTCGGGCAC	2132
Qy	690	SerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLysSer	708
Db	2133	ACCTTTGGCCAAAGTCTCCACGGCTGACCAACACTTACAGATCGGCCCTGGGTGTGGCG	2192
Qy	709	AlaProGlySerSerTyProAlaValPheValValGlyThrIleGlyGlyValThrGly	728
Db	2193	TCA---GGCTCGAACTGG---AACCTGTATGCTTTCGGCAC-----GGCCGCTCAGG	2240
Qy	729	Ala-----TyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGln	746
Db	2241	GCTCGCTCTACGCCAGCTGGAGACAGCGGGCGCTCTCTGGACGGACATCCAGGGCTCC	2300
Qy	747	HisGlnTyrglyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgVal	766
Db	2301	GGCTTCGGCTCCATCGACAGACCAAGGTGCGCGGACCGCGGACACCCCGCGGCAAGTC	2360
Qy	767	TyrIleGlyThrAsnGlyArgGlyIleValTy-----GlyAspIleGlyGlyAlaPro	784
Db	2361	TACGTGGGACCAACGCGCGGGGCGTCTTTTACGCTCAGGGAACCGCTCGGGCGGCGCAC	2420
Qy	785	SerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerPro	804
Db	2421	GGCGGGACT-----TCCTCGTCTGACCAAGCAGACAGCAGCAGC	2453
Qy	805	SerProSerSerSerProSerProSerProSerProSerProSerProSerSerPro	824
Db	2454	AGCAGTACCTCTCCGCGACGCTCGAGCACACGCTGAGGTGCGAGCGTTGATCCACGACC	2513
Qy	825	SerProSerProSerProSerPro-SerProSerArgSerProSerProSerAlaSerPr	844
Db	2514	CGGGCTTCGAGGTGACTTGTCTGAGGACACAGCTCGGCCCGCGGTCCACCGGGGTCA	2573
Qy	844	oSerPro	846

```

Db      2574  GTCCCG 2580
|||||
RESULT 5
ADD24893
ID      ADD24893 standard; DNA; 3668 BP.
AC      ADD24893;
XX
XX
XX      15-JAN-2004 (first entry)
XX
XX      DNA encoding Xanthomonas campestris cellulase #1.
XX
XX      Directed genetic engineering; galactomannanase; reduced activity;
KW      enhanced activity; xanthan gum production; suspension stability;
KW      emulsion stability; temperature resistance; pseudoplasticity; amylase;
KW      cellulase; extracellular protease; intracellular protease;
KW      glucose dehydrogenase; gene; ds.
XX
XX      Xanthomonas campestris.
XX
XX      US2003036176-A1.
XX
XX      20-FEB-2003.
XX
XX      10-AUG-2001; 2001US-00927827.
XX
XX      28-MAR-2001; 2001US-0279493P.
XX
XX      (BOWE/) BOWER S G.
XX      (RAMS/) RAMSEIER T M.
XX
XX      Bower SG, Ramseier TW;
XX
XX      WPI; 2003-625389/59.
XX      P-PSDB; ADD24919.
XX
XX      New transformed cell or organism having reduced or enhanced activity of
PT      at least one protein, useful for producing xanthan gum, which are useful
PT      for providing formulations and properties.
XX
XX      Claim 15; SEQ ID NO 21; 135pp; English.
XX
XX      The present invention relates to polypeptide and polynucleotide sequences
CC      from Xanthomonas campestris which may be used for activity reduction or
CC      enhancement using directed genetic engineering. A transformed cell or
CC      organism having reduced or enhanced activity of at least one such protein
CC      e.g. galactomannanase can be generated by disrupting the gene encoding
CC      the protein. The activity of the protein is reduced by the presence of an
CC      antisense nucleic acid sequence. The nucleic acid sequence of the gene
CC      encoding the protein is a recombinant sequence having at least one
CC      mutation as compared to the wild-type gene encoding the protein. The
CC      transgenic cell or microorganism are useful for producing xanthan gum,
CC      which are useful for providing formulations and properties, such as long-
CC      term suspension and emulsion stability in alkaline, acid, and salt
CC      solutions, temperature resistance, and pseudoplasticity. The present
CC      sequence encodes an enzyme relating to the present invention.
XX
XX      Sequence 3668 BP; 570 A; 1181 C; 1292 G; 625 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 7e-33 Length: 3668
Score: 1167.50
Percent Similarity: 48.55% Matches: 299
Best Local Similarity: 33.30% Conservative: 137
Query Match: 22.74% Mismatches: 351
DB: 9 Indels: 112
Gaps: 28

US-09-917-376-1 (1-957) x ADD24893 (1-3668)
Qy      23 ThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaThrAlaSer 42
|||
Db      1136 ACCAACAGGAGCATGGCTGGCTGCTTGGAGCCTGCTG---CTGTGTCTCTTTTCGACG 1192
|||||

```



399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAnProPro 418  
2231 GCGGTGTCATCACAGCGCGCGCGTGG-ACCGCACATGCGCGCGAC----- 2281  
419 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAnSer 438  
2282 -----TGGATGGG-----GCGCTGGCGATCGATCCGTTGACGGC 2317  
439 AspArgMetLeuTyrGlyThrGlyAlaThrLeuThrAlaThrAsnAspLeuThrLysTrp 458  
2318 AACCATGCGTGTCTGACCGCTACCGCATCTGGCGCTCGGCAATCTG----- 2368  
459 AspSerGlyGlyGlnIleHisIleAlaPro-----MetVal 470  
2369 -----CAGGATTTGCGCGCACCGCGCGCTGCGTGGTGTTCAGGAC 2416  
471 LysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeu 490  
2417 CGTGGGCTGGAAACCGCGCGCTGGACCTGCTCAGCCCGATGCTGGCGCGCATCTG 2476  
491 IleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAlaValProSer 510  
2477 CTCAGCGCTCGCGCATATCGAGCTTCGCGCATGACGACCTGGACCGCTG---CAG 2533  
511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 530  
2534 TTGCACTACCGCGCGCGCGCTGACCAATGGCGAAAGCATCGATGCGCGCGCAGCG 2593  
531 ProSerIleValArgAlaGlySerPhe-AppProSerSerGlnProAnAspArgHi 550  
2594 CCGCAGTGGTGTGGCGAGCGTACCGTGGCGCGCGCGCGCGCAACAGAAATCCGCGCG 2653  
550 sValAlaPheSerThrAspGlyGlyLysAsnTrpPheGln---GlySerGluProGlyG1 569  
2654 C---TGTATTACCGGTGTCGCGCGCGCAAGCAATGAGCTGCGTTGCGAGTGGCGCGCAGC 2710  
569 yValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPr 589  
2711 CGGGCAGGGCGCGCGCAGCATTCCTATGTCGCGATGCGCGCGCAGGTGTGTGGCACC 2770  
589 oGlyAspProGlyGlnProValValThrAlaValGlyPheGlyAsnSerTrpAlaAlaSe 609  
2771 GCGCGCGCGCGGCAAT-----TGGCGCACGTCGCGACTTCGCGCGCGCAGTGGCGCGTGT 2824  
609 rGlnGlyValProAlaAsnAlaGlnIleArgSerAspAtqValAsnProLysThrPheTy 629  
2825 GAACGCGCTCCCGAACCCCGCGTGGGTGTCGATCGGTGGAGCCAGCCGCGCTGGTA 2884  
629 xAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThrPheG1 647  
2885 TGGGTGGATGTCGCCAGCGCGCAGCTGTACGAGAGCAGCAGCATGCCGCGCGCAGTTCCG 2944  
647 nProValAlaAlaGlyLeu----- 653  
2945 T-----GCGACCGGTGTACAGTGGCGCAGCCCGCGCGCGATCAGCGCACCGCGCGCA 2998  
654 -----ProSerSerClyAlaValGlyValMetPheHisAlaValProGlyLysGluG1 671  
2999 GCTGCGTCCGAGACCGCGCGCGCGCGTGGTGTATCTGCGCAGCCCGCGCGCAAG----- 3053  
671 yAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTr 691  
3054 -----GGTGTGTCGCTGGCAGACGCT-----GGCT 3082  
691 pSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProG1 711  
3083 GCAGGTGCTCTCACAGCGCGCACCAAGCAGCTGCTGGCGCATCGGCAAGCGCTGGCGGC 3142  
711 ySerSerTyrProAlaValPheValGlyThrIleGlyValThrGlyAlaTyrAr 731  
3143 CGGTGGCGCGCGCGCGCTGTATCTGGCGCGCGCGTGGCGCGTGGATCGGTGTCCG 3202  
731 gSerAspAspCysGlyThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAs 751

3203 CTCGAGATGGCGCGCGCGCAGTGGCAGCGCATCAACGATGACGCGCAGCTTCGGCGG 3262  
751 nTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAs 771  
3263 C---CCGTACAGCGTACCGGTGATCCGCGATTGCCGGCGTGTGTACTTCCGCCACCGG 3319  
771 nGlyArgGlyIleValTyrGlyAspIleGlyAlaProSerGlySerProSerProSe 791  
3320 CGCGCGCGCATTTCTACGGCGATCCGAGATGAGCAGTGCCTGCTGCGCGA----- 3374  
791 rValSerProSerAlaSerProSerLeuSerProSerProSerProSerProSerPro-- 810  
3375 ---CGTCTGCTGCTTTACCGTACCGCGCGCGCGCTGGCGCGTGGCGCGCGCGT 3430  
811 -----SerProSerProSerProSerProSerProSerProSerProSerProSer-- 827  
3431 ACACCGCGGACCTGGAGATCACCGCATGACCCCGT-CCACATTTCCGCGCGCTGCTTG 3489  
828 -----ProSerProSerProSerProSerArg----- 836  
3490 CTGGGTGTGTGTGGCGCAGCGCGCGCGCAACCGAGCGCTGCGGTGCTGTT 3549  
837 -----SerProSerProSerAlaSerProSerProSerProSerProSerSerPr 850  
3550 GCCGATGGCGGTGCTGCAGCGCGCATGAGCAATGCCGCTATGGGGTGGCGCGCACGCC 3609  
850 oSerProSerSer-----SerProSerSerProSerProSerProSerPro 864  
3610 GCGCGCGCAATCAGGTGAGTTTCATGGCAAGCAGGCCACCGCCACAGACCG 3661

## RESULT 6

ADD42054  
ID ADD42054 standard; DNA; 5698 BP.  
XX AC ADD42054;  
XX DT 15-JAN-2004 (first entry)  
XX DE Trichoderma reesei Family 74 xyloglucanase-encoding DNA, SEQ ID NO:1 #1.  
XX KW Trichoderma reesei; Family 74 xyloglucanase; beta-1,4-glycosidic linkage;  
XX KW xyloglucan hydrolysis; biomass degradation; cellulose; hemicellulose;  
XX KW ethanol production; detergent composition; fabric treatment;  
XX OS Hypocrea jecorina.  
XX PN WO2003089598-A2.  
XX PD 30-OCT-2003.  
XX PF 17-APR-2003; 2003WO-US011831.  
XX PR 19-APR-2002; 2002US-0373987P.  
XX PA (NOVO ) NOVOZYMES BIOTECH INC.  
XX PI Michael R, Zaretsky E, Haas J;  
XX DR WPI; 2003-845528/78.  
XX DR P-PSDB; ADD42055.  
XX PT New polypeptides having Family 74 xyloglucanase activity, and encoding  
XX PT nucleic acid molecules, useful for degrading cellulose- and hemicellulose  
XX PT -containing biomass to ethanol or as a detergent.  
XX PS Claim 1; SEQ ID NO 1; 96pp; English.  
XX CC The invention relates to a Family 74 xyloglucanase (ADD42055) from the  
XX CC fungus Trichoderma reesei (Hypocrea jecorina), and nucleic acids encoding  
XX CC it (e.g., ADD42054). The enzyme hydrolyses the beta-1,4-glycosidic  
XX CC linkages in the backbone of xyloglucan to xyloglucan oligosaccharides.



CC The invention also relates polypeptide sequences at least 70% identical  
 CC to the enzyme, expression vectors and host cells comprising a nucleic  
 CC acid of the invention, the recombinant production of the enzyme, and  
 CC mutant enzymes and the nucleic acids encoding them. The xyloglucanase of  
 CC the invention can be used in the degradation of cellulose- and  
 CC hemicellulose-containing biomass to produce ethanol. It can also be used  
 CC in a detergent composition for treating fabric during a machine washing  
 CC cycle. The present sequence represents a trichoderma reesei Family 74  
 CC xyloglucanase-encoding DNA. Note: The present sequence (referred to as  
 CC SEQ ID NO:1) is given as shown in the sequence listing. A comparison of  
 CC this sequence with that shown in figure 1 (also referred to as SEQ ID  
 CC NO:1) shows that the present sequence is twice as long as the figure 1  
 CC sequence (ADD42060) and contains alternate 60 bp blocks of sense strand  
 CC and complementary strand sequence.

XX  
 SQ Sequence 5698 BP; 1145 A; 1704 C; 1704 G; 1145 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 1.04e-24 Length: 5698  
 Score: 940.00 Matches: 374  
 Percent Similarity: 30.11% Conservative: 136  
 Best Local Similarity: 22.08% Mismatches: 315  
 Query Match: 18.31% Indels: 871  
 DB: 9 Gaps: 59

US-09-917-376-1 (1-957) x ADD42054 (1-5698)

QY 25 SerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44  
 DB 128 TCTCGAGTCTTCCCTTGCCTGGGGGCGGTC-----ATCCCTGCC 169  
 QY 45 HisAlaAlaThrThrGln-----GluGlyAlaPro----- 50  
 DB 170 CATGCTGCTTCTTCGAGAGCTCAGGAACGGGAACAGACCCTCCGGCAGTAGGACGG 229  
 QY 51 -----ProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGly 67  
 DB 230 GTACGCGGAATTCATGGAAGACGTCAAGTCTCGGCGGCGGCGGCTTCGTCGCCGCGC 289  
 QY 68 IleValPheAsn-----GluGlyAlaPro----- 75  
 DB 290 ATCATCTTCCAAAGTACCTTTTCGAGTTCGAGCGCGCGCGCGCGGACGAGCGGCGCG 349  
 QY 76 -----GlyIleLeuTyrValArgThrAspIleGlyGlyMet 87  
 DB 350 TAGTAGAGGTTCCGACAGAAAGGGGTAGCATATGCACAGATATTTGGCGGCTG 409  
 QY 88 TyrArgTrp----- 90  
 DB 410 TACCGCTCAAGGGTTCTGTTTCCGCATCGTATACGTGTTCTCTATTAACCGCCGAC 469  
 QY 91 -----AspAlaIleAlaGlyArgTrpIleProLeuLeuAspTrpVal----- 104  
 DB 470 ATGCGCGAGTTCCGCGACGACTCATGACCGCGCGTCAACGATGGGATTTGCTGTAATGCC 529  
 QY 105 GlyTrp-----AsnAsnTrpGly----- 110  
 DB 530 GGCTGGCACAAGCGGTGTGAGTACTCTGGCGGAGTGCCCTACCTTAACGACTATTACGG 589  
 QY 111 -----TyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIleVal 127  
 DB 590 CCGACCGTGTTCGGGGCATCGAGCTGTTCGCTTGATCCGACGACGATCAAAAGGTG 649  
 QY 128 TrpAlaAlaVal----- 131  
 DB 650 TATGCCGAGTACCCCGTAGCTGCGACACACGGAACCTAGCGCGTCTGCTAGTTTCCAC 709  
 QY 132 -----GlyMetTyrThrAsnSerTrpAsp----- 139  
 DB 710 ATACGGCGTTCACGCGCATGTATACGAACAGCTGTC-TGTGATGTCCTCAGATCTAGACC 768  
 QY 139 ----- 139

DB 769 TATGATTGGACGGCGTACATATGCTTCTCGACCAGACACTACAGAGAGTCTAGATCTGG 828  
 QY 140 -----ProAsnAspGlyAlaIleLeuArg 147  
 DB 829 ATACTAACCTCGCTGACATTGGCCATATATAGGGATCCGAGTAATGGAGCCATCAATCGC 888  
 QY 148 SerSerAspGln----- 151  
 DB 889 TCGTCAGACCGCGACTGTAAACCGGTATATATCCCTAGGCTCATTAACCTCGGTAGTAGCG 948  
 QY 152 -----GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsn 167  
 DB 949 AGCAGTCTGGCGGCGCAACGTTGCTTCCACCAACTTGCCTTCAAGTTCGGGGGTAAAC 1008  
 QY 168 MetProGlyArg----- 171  
 DB 1009 ATGCCAGACCGCGGTTGCACACAGGAAGTGGTTGAACGGGAGTTTTCAGCCCCCATTTG 1068  
 QY 172 -----GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeu 187  
 DB 1069 TACGCTCTCGGGGAGCGGAGAGCGTCTGGCTGTGATCCGCGCAACTCCAAACATCATC 1128  
 QY 188 TyrPheGlyAlaPro----- 192  
 DB 1129 TACTTTGGTGTCTCTCGGCTCTCCACAGACCGACAGTAGGCCGTTGAGGTTGTAGTAG 1188  
 QY 193 -----SerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrp 207  
 DB 1189 ATGAACACACAGCGCTCAGAAACGCGCTCTGGAAAGTCTACGAGCGCGCGGTGACCTTT 1248  
 QY 208 SerGlnMet----- 210  
 DB 1249 TCCAGGTTCTCGCGAGTCTTTGCGGAGACCTTTCAGATGCTGTCGCGCGCACTGGAAA 1308  
 QY 211 -----ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThr 227  
 DB 1309 AGGTTCCAGAGCTGTTTTCACGCAACTGGGAGCTATATCCAGACCCGAGTGTATCCAAC 1368  
 QY 228 GlyTyrGln----- 230  
 DB 1369 GGCTAACACAGCAGCAAGTGGCTGACCTGCATGTAGGTTCTGGGCTCATCAAGGTTG 1428  
 QY 231 -----SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSer 247  
 DB 1429 CCGATGTTGTCGACACAGCAAGACTCATGTGGTTCAGTTCGACTCAACACGACGACG 1488  
 QY 248 LeuGlyGlnAla----- 251  
 DB 1489 ACCGGGGAGCGCTGTTCTGTTCTGAGTACACCCCAATGCAAGCTGAGTTGGTCTGCTGC 1548  
 QY 252 -----SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPhe 267  
 DB 1549 TGGCCCCCTCGAGCGTCTGATCTTTGTTGGCAGCGGTGATAACATCACTGCTTCAGTC 1608  
 QY 268 Trp----- 268  
 DB 1609 TATGTGACACGTGACGAGCATAGAAACACCGTCCGCACTATTGTAGTGACGAAGTCAG 1668  
 QY 269 -----SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286  
 DB 1669 ATACACTCGTGAATGCGGCTCCACGTTGGAGTGTGTGTACCGGGGCGAGCCAGGAAATAC 1728  
 QY 287 IleProHis----- 289  
 DB 1729 TTTCTCACAAGTTACGGCCGAGGTGCACCTCACGACATGGCCCGTCCGTTCCCTTTATG 1788  
 QY 290 LysGlyValPhe-----AspProValAsnHisValLeuTyrIleAlaThrSerAsn 306  
 DB 1789 AAAGAGGTTTCGGGAAACTGCAGCCGACGAGGAAGGCGCTTGTATCTGACCTATTCCGAT 1848  
 QY 307 ThrGlyGlyPro----- 310  
 DB 1849 GGCACAGGGCGCGCTTTGACGTGCTCTTCCGGNAACATAGACTGGATAAGGCTA 1908

Qy 311 -----TyrAspGlySerSerGlyAspValTrpIysPheSerValThrSerGly 326  
Db 1909 CCGTGTCCCGGCTATCATGGCACACTTGGCTCAGTGTGGAGTACGACATTGCGAGGGGA 1968  
Qy 327 ThrTrpThrArgIleSerProValProSerThrAspThr----- 339  
Db 1969 ACTTGGAAAGACATACTACCGTGTGAACCGAGTACACACTCCATGTGTAAACGTCCCCCT 2028  
Qy 340 -----AlaAsnAspTyrPheGlyTyrSerGlyLeu 349  
Db 2029 TGAACCTTTCTGATCACCCCTGTCTCTGGATCAGATCTATCTTTGGCTTTGGCGGCTT 2088  
Qy 350 ThrIleAsp----- 352  
Db 2089 GGCCTCGATTGTAGTGGGACAGAGACCTAGTCTAGATATGAACCGAAACCGCGGAA 2148  
Qy 353 -----ArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpPro 369  
Db 2149 CCGAGCTAAACCAAAAGCCAGGAACCTTGTGTGCTTCTTTGAACTCTTGTGGCCA 2208  
Qy 370 AspThr----- 371  
Db 2209 GATGCTCAGCTGTTTGGTCTCTTGGGAACAACAAGAACTTGAACAACCCGCT 2268  
Qy 372 -----IleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrp 389  
Db 2269 CTACGAGTCGACTTCTCGTGCACCGACTCTGGGACACATGGAGCCGATCTGGCGGTG 2328  
Qy 390 ThrSerTyrProAsnArgSer-----LeuArgTyrValLeuAspIleSerAlaGluPro--- 407  
Db 2329 GCGAGCTATCCGAAAGCCAGCTGGTGTAGACC-CTGTTGTACCTCGGGCTAGACCCGAC 2387  
Qy 407 ----- 407  
Db 2388 CCGCTCGATAGGCACTGAGACCTATTACTACAGCATCTCAGTAGTCACCTTTAACGATC 2447  
Qy 408 -----TrpLeuThrPhe-----GlyValGlnProAsn----- 416  
Db 2448 CGATGCGAATGGCTGACTCTGGATAATGATGTCCTAGAGTCACCTCAGTGAGATGTGTAG 2507  
Qy 417 -----ProProValPro----- 420  
Db 2508 GCTACGCTTACCGACTGACTCTGCTTCAAGACTCCCAAAAGCACCGTGGATCAAGAAACAAC 2567  
Qy 420 ----- 420  
Db 2568 TTTATCGATGTGATGACTGAGACGAAGTTCAGGGTTTCGTGGCACTAGTTCTTGTG 2627  
Qy 421 -----SerPro-----LysLeuGlyTrpMet 427  
Db 2628 AATAGCTACACTCGAGCGAGTCACCGTCCGATGGTCTCATCAAGCGCTCGGCTGGATG 2687  
Qy 428 AspGluAlaMetAla----- 432  
Db 2688 ATTGAGTCTCTCGGCTCGCTCAGTGGCAGGCTACAGAGTAGTTTCGGGAGCGGACCTAC 2747  
Qy 433 -----IleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAla 447  
Db 2748 TAACTCAGAGAGCAGATTGACCAACCGACAGCAACCACTGGCTCTACCGCACCGGAATG 2807  
Qy 448 ThrLeuTyrAlaThr----- 452  
Db 2808 ACAATCTTTGGCGCTTAACTGGGTTGGCTGTCTGTGTGTGACCGAGATGCCGTGGCCTTAC 2867  
Qy 453 -----AsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAla 467  
Db 2868 TGTTAGAAACCGCGCCACGATCTCACCAACTGGGACAGCGCCACATGTGTCAATCAA 2927  
Qy 468 ProMetValLysGly----- 472  
Db 2928 TCACTGGCAGACCGCGTGTAGAGTGGTTGACCCCTGTGCGGGTGTATTACACAGTTAGTT 2987

Qy 473 -----LeuGluGluThrAlaValAsnAspLeuIleSerProSerGly 487  
Db 2988 AGTGACCGTCTGCCCATCGAGGAATTCTCCGTCCAGGACCTGGCTCGACACCGCGGA 3047  
Qy 488 AlaProLeuIleSer----- 492  
Db 3048 AGCGAGCTATTGGCTAGTCTCTTAAGAGGCGAGTCTTGGACCGGAGAGCTGGGCGGCT 3107  
Qy 493 -----AlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAla 507  
Db 3108 TCGCTCGATAAACCCCGCAGTCGGAGACGACAACGGCTTACCTTTGCCACGAGAAACGAC 3167  
Qy 508 Val----- 508  
Db 3168 CTGGGACATCGCGCGCTCAGCTCTCTGTGTCCGAAAGTGGAAACGGTCTGTTTGTG 3227  
Qy 509 -----ProSerThrIlePheThrSerProValPheThrThrGlyThrSerVal 524  
Db 3228 GAGCCCTGTAGCGGCGAGAGCTCTGGCAACGCCCATATGGGCCACTCTCGACGAGCGTC 3287  
Qy 525 AspTyrAla----- 527  
Db 3288 GACTACCGCGGAGCGTCTGCCAGACCCGTTGGGGTGTACCCGGTGGAGCTGCTCGCAG 3347  
Qy 528 -----GluLeuAsnProSerIleIleValArgAlaGlySer----- 539  
Db 3348 CTGATCGCGGCTACTCGGTCAAGAGCGTCTCGCGCTCGGCAACACCGCGGACGCA 3407  
Qy 540 -----PheAspProSerSerGlnProAsnAspArgHisValAla 552  
Db 3408 CAGGTGGCCATTTGAGCCAGTCTTCGACGAGGCGCAGCCGTTGTGGCGCGCGTGTG 3467  
Qy 553 Phe-----SerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 569  
Db 3468 GTCCACCGGTAGACGTCCGACGGCGCGCGAGTGGAGCATCGACTACGCGGCCACACG 3527  
Qy 570 ValThrThrGlyGly----- 574  
Db 3528 TCCATGAACGCGCGCAGGCTCCGCGCGCTGCACCTCGTAGCTGATCGCGCGGTGTC 3587  
Qy 575 -----ThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 589  
Db 3588 AGGTACTTGGCGCACCGTGGCTATTTCGCGCGACGCGCACGATCTCTGGTCGACC 3647  
Qy 590 GlyAspProGly----- 593  
Db 3648 GCCTCTCGCGCGGTGCCACCGGATAGCCGGTGCCTGTGTGTAGGAGACCGAGTGTG 3707  
Qy 594 -----GlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGly 611  
Db 3708 CGGAGCAGCGCGCTGCAGCGCTCG--CAGTTCCAGGCGCAGCTTTGCTCTCGTCTCGAGC 3764  
Qy 612 ValProAlaAlaGln----- 617  
Db 3765 CTGCGCGCGCGCGCGAGCTCGCGAGCGTCAAGTCCCGTCCGAAACGGAGGAGAGCTCG 3824  
Qy 618 -----IleArgSerAspArgValAsnProLysThrPheTyrAlaLeu 631  
Db 3825 GACGGCGCGCGCGCTCATCGCTCGGACAAAGAACCAACAGCGCTCTTACCGCGCGC 3884  
Qy 632 SerAsnGlyThrPhe----- 636  
Db 3885 TCCGATCGACCTTTAGTAGCGGAGCGCTTCTTCTGTGTTCTCGCAAGATGCGGCGC 3944  
Qy 637 -----TyrArgSerThrAspGlyValThrPheGlnProValAlaAla 651  
Db 3945 AGGCTAGCTGAAAAACGTCAGCAAGGACACCGCGCAGCAGCTTC-----ACGCGC 3995  
Qy 652 GlyLeuProSerSerGlyAlaValGlyValMet----- 662  
Db 3996 GGG---CCCAAGCTGGGCGCGCAGGGATGAGTCTTCTGTGCGCGCTGTCGAGTGC 4052  
Qy 663 -----PheHisAlaValProGlyLys 669







Db 1354 TCGTCGCCGCTGTCGGTGGGTCGAGGTGCAGTACAGAACAAATGATTCGGCGCCGGGT 1413  
QY 886 AspAsnGlnIleIysProGlyLeuGlnValValAsnThrGlySerSerValAspLeu 905  
Db 1414 GATACACAGATCAACCGGGTCTCGAGTGGTGAATACGGGTCGTCGGTGGATTG 1473  
QY 906 SerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerThrLeuValTyr 925  
Db 1474 TCGACGGTCACGGTCGGTACTGCTGTCACCGGGATGGTGGTTCGTCGACACTGGGTGAC 1533  
QY 926 AsnCysAspTrpAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 945  
Db 1534 AACTGTGACTGGCGCGGATGGGGTGGGAATATCCGGCCCTCGTTCGGCTCGGTGAAC 1593  
QY 946 ProAlaThrProThrAlaAspThrTyrLeuGln 956  
Db 1594 CCGCGACCGCGGACGGGACACTACTGTCAG 1626  
RESULT 9  
ID ABZ76162 standard; DNA; 3365 BP.  
XX AC ABZ76162;  
XX DT 29-MAY-2003 (first entry)  
XX DE A. cellulolyticus Gux1 protein encoding DNA.  
XX KW Gux1; thermal tolerant; exoglucanase; glycoside hydrolase; cellulose;  
KW biofuel; detergent; pulp; paper processing; feed processing; textile;  
XX cellullase; gene; ds.  
XX OS Acidothermus cellulolyticus.  
XX FH Key Location/Qualifiers  
FT CDS 1..3365  
FT FT /\*tag= a  
FT FT /product= "Gux1"  
FT FT /transl\_except= (pos: 682..683, aa: Pro)  
FT FT /notes= "this codon has an apparent one nucleotide  
FT FT basepair deletion which alters the reading frame"  
XX WO2003012095-A1.  
XX PN 13-FEB-2003.  
XX PD 28-JUL-2001; 2001WO-US023820.  
XX PF 28-JUL-2001; 2001WO-US023820.  
XX PR (WIDE ) MIDWEST RES INST.  
XX PA Adney WS, Ding S, Vinzant TB, Himmel ME, Decker SR;  
XX PI Lantz McCarter S;  
XX PI WPI; 2003-300494/29.  
XX DR P-PSDB; ABP71656.  
XX XX New thermal tolerant Gux1 peptide having specified amino acid sequence,  
XX XX useful in the degradation of cellulose to biofuels.  
XX PT Disclosure; Page 22-23; 44pp; English.  
XX PS The invention relates to a thermal tolerant Gux1 peptide from A.  
XX CC cellulolyticus. The Gux1 exoglucanase is a member of the glycoside  
XX CC hydrolase family and comprises a catalytic domain GH48, carbohydrate  
XX CC binding domain type III, and a carbohydrate binding domain type II. The  
XX CC polypeptide is useful in the degradation of cellulose into biofuel, or  
XX CC for conversion of biomass to biofuel additives. It is used in detergents,  
XX CC pulp and paper processing, food and feed processing, and in textile  
XX CC processing. It can also be used alone or in combination with other  
XX CC cellulase or glycoside hydrolases. The novel polypeptide generates

CC alternative cellulase enzymes capable of assisting in the commercial-  
CC scale processing of cellulose to sugar for use in biofuel production. The  
CC present sequence represents a A. cellulolyticus Gux1 cellulase encoding  
CC DNA  
XQ Sequence 3365 BP; 669 A; 1117 C; 1041 G; 538 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 6.68e-07 Length: 3365  
Score: 429.00 Matches: 79  
Percent Similarity: 97.62% Conservative: 3  
Best Local Similarity: 94.05% Mismatches: 2  
Query Match: 8.35% Indels: 0  
DB: 7 Gaps: 0  
US-09-917-376-1 (1-957) x ABZ76162 (1-3365)  
QY 873 ValIysValGlnTyrIlyAsnAsnAspSerAlaProGlyAspAsnGlnIleIysProGly 892  
Db 112 CTCAAACCGCAGTATAAGAACATGATTCGGCGCGGAGTACCAACAGATCAACCGGGT 171  
QY 893 LeuGlnValValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 912  
Db 172 CTCAGTTGGTGAATACCGGGTCGTCGTCGGTGGATTGTCGACGGTGCAGTGGGTAC 231  
QY 913 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaIle 932  
Db 232 TGGTTTACCCCGGATGGTGGTTCGTCGACACTGGTGTACAACTGTGACTGGCGCGGATG 291  
QY 933 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 952  
Db 292 GGGTGTGGGAATATCCGGCCCTCGTTCGGTTCGGTGAACCCCGGACGCCGCGCGGAC 351  
QY 953 ThrTyrLeuGln 956  
Db 352 ACCTACTCTGCAG 363  
RESULT 10  
AAX55661  
ID AAX55661 standard; DNA; 11706 BP.  
XX AC AAX55661;  
XX DT 30-JUL-1999 (first entry)  
XX DE DNA sequence encoding truncated cellulases Cel B4/5 and Cel B5.  
XX KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;  
KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;  
XX cotton-containing fabric; stonewashing; ss.  
XX OS Unidentified.  
XX FN EP921188-A2.  
XX PD 09-JUN-1999.  
XX PF 15-SEP-1998; 98EP-00810919.  
XX PR 19-SEP-1997; 97US-00932571.  
XX PA (CLRN ) CLARIANT FINANCE BVI LTD.  
XX PI Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;  
XX PI Morgan H, Williams DP;  
XX DR WPI; 1999-315403/27.  
XX DR P-PSDB; AAY13492.  
XX XX New truncated cellulase proteins, useful in detergents and for producing  
XX XX 'stonewashed' denim.  
XX PS Disclosure; Page 20-23; 65pp; English.

XX The invention relates to a recombinant cellulase active protein free of  
 CC proteinases of native thermophilic and alkaliphilic origin, comprising  
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel E1/2/3,  
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-  
 CC length sequences, or functional equivalents. Cel B5 extends from amino  
 CC acid A1011 to F1424 or K1425 or N1426, and Cel B4/5 extends from amino  
 CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from  
 CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3  
 CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751  
 CC and the stability region extends from amino acid E482 to G635 in the  
 CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new  
 CC enzymes are useful in laundry detergent compositions to prevent or remove  
 CC staining, backstaining or graying, for use on cellulosic materials  
 CC including cotton-containing fabrics. They are especially useful for  
 CC preventing redeposition of colorant during stonewashing, and for  
 CC processing of textiles where cellulose breakdown is required. The new  
 CC truncated enzymes show reduced redeposition of dye compared to using non-  
 CC truncated cellulase compositions  
 XX  
 SQ Sequence 11706 BP; 3828 A; 1994 C; 2994 G; 2890 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 9.59e-06 Length: 11706  
 Score: 409.00 Matches: 263  
 Percent Similarity: 32.37% Conservative: 163  
 Best Local Similarity: 19.98% Mismatches: 386  
 Query Match: 7.96% Indels: 504  
 DB: 2 Gaps: 58

US-09-917-376-1 (1-957) x AAX55661 (1-11706)

QY 14 ArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGly 33  
 DB 735 AAGAGGGAATTCATCTTTTTCCTTTATGTTTATTAATAAACCGCTGTAGGT 791  
 QY 34 ValLeuProIleAlaIleThrAlaSerProAlaHieAlaAlaThrGlnProTyrThr 53  
 DB 792 ACTTGG-----ATATTTTCATCAGGAGCAAAAGCAGCAGCA-----TATACT 833  
 QY 54 TrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGly 73  
 DB 834 -----GTTGATTTTGAAGT 848  
 QY 74 AlaProGlyIleLeuTyr-----ValArgThrAspIleGlyGly 86  
 DB 849 GCTGATACCTTTATCTTACTTTGCTTATGGAATTCGAGCATAGCAGTTGACATGGGCAAT 908  
 QY 87 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTyrValGlyTyr 106  
 DB 908 ----- 908  
 QY 107 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 126  
 DB 909 -----GCATATAATGCTAAAGTAGTGTCTAGGCTGTCA-----AATAGAAGTTCA 953  
 QY 127 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146  
 DB 954 ATATGGGATGGAGTTGCAGTT-----GACGTTTAAA 983  
 QY 147 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThr----- 158  
 DB 984 AACATTATGAACATGAACCAATGCTGGTAGTTTCAGCGGTATGTAAACATAGCTACCAG 1043  
 QY 159 ---ProLeuProPheLysLeu-----GlyGlyAsnMetProGlyArg 171  
 DB 1044 AAGCCGGTTCATTTGGTATCTCAGCGGTTTACGACGATGGAAGTGGGTTAAGACTACT 1103  
 QY 172 GlyMetGlyGlnArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 191  
 DB 1104 CTCATAGGTGAGGTGTGGCTATT-----CCAAAT----- 1133  
 QY 192 ProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 211

DB 1134 -----TATGGAGAAATTT----- 1148  
 QY 212 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 231  
 DB 1149 -----GTTGGTAATGGACTCAATATTAGCAATGTCCAGGAATTTGTTAATT 1196  
 QY 232 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 251  
 DB 1197 GTAATACACACAAATGTAGAAAGCGCAAGTAGATATAATGTTGAC----- 1241  
 QY 252 SerLysThrIlePheValGlyValAlaAspProAsn-----AsnProValPhe 267  
 DB 1242 -----TATATCCAAATATGATGATATACTTACCTATCAATACGAGTGACA 1289  
 QY 268 TrpSerArg-----AspGlyGlyAlaThr-----TrpGlnAla-----Val 279  
 DB 1290 TTTTCAAGTGGATTGAAAGTGGCACTACCGAGGTTGCGAGGCAAGGGAAGCGGTGT 1349  
 QY 280 ProGlyAlaProThrGlyPheIleProHieLysGlyValPheAspProValAsnHisVal 299  
 DB 1350 ACAGTAAACCCAGATAGCGTTGTGCATATAGTGGCAAGTATAGT----- 1394  
 QY 300 LeuTyrIleAla-----ThrSerAsnThrGlyGly-----ProTyrAspGly----- 313  
 DB 1395 TTGTACGTCAGTGGGAAGAACCGTCAAAATGGCATGGTGCACAGATTCGCGTAGATACAATT 1454  
 QY 314 ---SerSerGlyAspValTrpLysPheSerVal-----ThrSerGlyThrTrp 328  
 DB 1455 TTGAAACACGGGTAAAGTGTATAAATAAGTTTGGGTTTATCAGAACAGTGGTCAACT 1514  
 QY 329 ThrArgIleSerProValProSerThrAspThrAlaAsnAsp---TyrPheGlyTyrSer 347  
 DB 1515 CAAAAATGTCATTAACCTATCCAAAGAGATTGTCTACAGATCCTTCAACAAGCTATGAA 1574  
 QY 348 GlyLeuThrIleAspArgGlnHieProAsnThrIleMetValAlaThrGlnIleSerTrp 367  
 DB 1575 AATCTGATATATAACAGGAGTGTACCGAGTAATACTGGTGGTGGTGGTGGAGTGAAGCTAC 1634  
 QY 368 TrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp 387  
 DB 1635 -----TCAATTCCTGCTGGTGTACAGTTAGCAGGTTGTTG 1670  
 QY 388 AspTrpThrSerTyrProAsnArgSerLeuArgTyr-----ValLeu 401  
 DB 1671 CTTTATGTTGAGGCACAAAATGCAAAATTTGGCTTCTCGGTTGATGATTTAAGATTAT 1730  
 QY 402 AspIleSer-----AlaGluPro----- 407  
 DB 1731 GATTTATCCAAAGTTGGCTGAACCTGAATGGGAGATACCCTCTTTCATAGAAAAGTAGA 1790  
 QY 407 ----- 407  
 DB 1791 GATTATTTCAAAGTAGGAGTAGCTTTGTCTTACAAAAGCATTTGCCTCTGATACAGAAAAG 1850  
 QY 407 ----- 407  
 DB 1851 AAGATGTTTTGAAGCATTTCAATAGTATTTCTGAGGGAACGAATGAACCATCAGAG 1910  
 QY 408 -----TripleThrPheGly 412  
 DB 1911 TTACTTGTGTAAGAAATACTTACAACTTTAGCAAGCAGACGAATTTGTAATTTTGCA 1970  
 QY 413 ValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAla----- 430  
 DB 1971 ACAAGTAACAACATTTGCCATCAGAGGTACATACCTGTTTGGCATGAGCAACACCCGAC 2030  
 QY 430 ----- 430  
 DB 2031 TGGTTTTTCAAGGACACAAATGGAATACGTTGAGCAAGGATGCATTGCTAAGCAGATTA 2090  
 QY 430 ----- 430



Db 2091 AAACAGTATATTATACGGTAGTGGGAAGATATAAAGGGAAGGTTTATGATGGGATGTG 2150  
Qy 431 -----MetAlaIleAspProPheAsnSerAspArgMetLeu----- 442  
Db 2151 GTAAATGAAGCAATAGATGAAGTCAAGGTGATGGATTCAGGAGATCTAACTGGTACAAC 2210  
Qy 443 -----TyrGlyThrGlyAlaThrLeuTyrAla----- 451  
Db 2211 ATTTGTAGTCCGGAATATATTGAGAAGGCTTTTATATGGGCACATGAAGCCGATCCAGAC 2270  
Qy 452 -----ThrAsnAspLeuThrLysTrpAspSerGlyGlnIleHisIleAla 467  
Db 2271 GCAAAATGTTTACCAACCATATACACACAGAAAACAGTCAGAGAGACAGTTTATTATAC 2330  
Qy 468 ProMetValIleSerGlyLeuGluGluThrAlaValAsnAspLeuIleSerProPheSerGly 487  
Db 2331 AACATGATTAAAGATCTCAAGGAAAAGGTGTT-----CCAATTCTATGGA 2375  
Qy 488 AlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAla 507  
Db 2376 ATAGGATTGCAGATCATATA-----AATCTTGATTGGCCCTCGATTAGCGAG 2423  
Qy 508 ValProSerThrIle-----PheThrSerProValPheThrThrGlyThrSerValAsp 525  
Db 2424 ATAGAGAACACCATAGATTGTTTCAGCTCT-----ATACCTGGATTGGAGATACAC 2474  
Qy 526 TyrAlaGluLeuAsnProSerIleIleValAlaGlySerPheAspProSerSerGln 545  
Db 2475 ATTACGAGCTTGATGATGATGTTTATCAGTGGGTTCCAGTACCAGTTACTCAACGCCA 2534  
Qy 546 ProAsnAsp----- 548  
Db 2535 CCAAGAGATCTCCTGATATAAACAGGCAATGAGATATAAGGAGTTATTGATTGTTTAA 2594  
Qy 549 -----ArgHisValAlaPhe----- 553  
Db 2595 AAGTACAACAATGTATAACAAGTGTAAACATTCTGGGAGCTGAAGGATGATTACTCATGG 2654  
Qy 554 ---SerThrAspGlyGlyLysAsn----- 560  
Db 2655 CTGAGTCAAACTTGGAAAAGTGAATGATTACCCGTTGTTATTGATGAAACTATAATCA 2714  
Qy 561 -----TrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAla 577  
Db 2715 AAATATGCTTTTGGAGCCTGATTGAGCCAACTGTGATACCGGCCAACTCAACATTCGCCA 2774  
Qy 578 Ala-----SerAlaAspGlySerArgPheVal 586  
Db 2775 GCACCAACCACTATTCAAAATACCTACACCAACTCCACACCAACCCCGACACGACAGTG 2834  
Qy 587 TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrp 606  
Db 2835 AGTGCAACGCCAACACAGCA-----CCGACGGCATCCCGTAGGT---GGCAGTTACTGG 2888  
Qy 607 AlaAlaSerGlnGlyValProAla-----AsnAlaGlnIle 618  
Db 2889 ACGCCGAGTGAGAGTTACAGTGGCTGAGGATGATGTTGCGAATGGAAATTAAGCAGC 2948  
Qy 619 ArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsn-----GlyThr----- 635  
Db 2949 CCGACCAATGATTGAATCTTAAG-----ATAAAGATAGAGAATGTTGGGACGACAGCG 3002  
Qy 636 -----PheTyrArgSerThrAspGlyGlyValThr----- 645  
Db 3003 GTAGATCTTAGCAGGTGAAGGTAAAGATACTGTTACACGATAGATGTTGAGGCAACACAG 3062  
Qy 646 -----PheGlnPro-----ValAlaAla 651  
Db 3063 AGTGTAAGTGTAAACAGCAGCATAGATCTTCGGTATATAGATGTGAAGTTGTGAAGCTT 3122  
Qy 652 GlyLeuProSerSerGlyAla-----ValGly 660  
Db 3123 GGAGCGAACCGAGCGGAGCGGATTACTATGTGGAGATAGGCTTTAAGAGTGGAGCAGGG 3182

Qy 661 ValMetPheHisAlaValProGlyLysGlyGluGlyAspLeuTrpLeuAlaAlaSerSerGly 680  
Db 3183 GTTTTGGCAGCAGGCGCAAGCAGCAAGAGAGATAGACTTAGCATACAGAAAGCGCAGTGC 3242  
Qy 681 LeuTyrHisSerThrAsnGlyGlySerSerTrpSerAla----- 693  
Db 3243 AGCTACAAATCAGTCAAAATGACTATTCCGTGAGGAGTGCACAGGCTATATAGAACGAG 3302  
Qy 694 ---IleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySer 712  
Db 3303 AAGTACACGGTATATAGATGATGTTCTATCGGGAAGAGACCGGAGCAGAACGCC 3362  
Qy 713 SerTyrProAlaValPheValGlyThrIleGlyValThrGlyAla 729  
Db 3363 CAGATCAAGGTATGTTGATCGAATGGAATTTAAGCAGCCCGCAATGTTATGTAATCCT 3422  
Qy 730 ---TyrArgSerAspAspCysGlyThrThr----- 738  
Db 3423 AAGATAAGATAGAGAAATGTTGGACGACACAGCGGTAGATCTTACGAGGTTGAAGTAAAG 3482  
Qy 739 ---TrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThr 757  
Db 3483 TACTGGTACCATAGATGTTGAGGCAACACAGAGTCTAAGTGTAAACAGCAGCATAAAC 3542  
Qy 758 GlyAspHisAlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIle---Val 776  
Db 3543 CCTCGGTATATAGATGTGAAGTTTGTGAAGCTTGGAGCAAAATGCAAGTGGAGCGGATTAC 3602  
Qy 777 TyrGlyAspIleGly-----GlyAla----- 783  
Db 3603 TATGTGAGATAGAGCTTTAAGAGTGGAGCGGGTTTGGCAGCAGGCGGAGCAGCAGAAAG 3662  
Qy 784 -----ProSerGlySer----- 787  
Db 3663 GAGATAAGACTTAGCATACAGAGGCGAGTGGCAGCTACAACTCAGTCAATGACTATTTCG 3722  
Qy 787 ----- 787  
Db 3723 GTGAGGAGTCAACAGGCTATATAGAGAAACGAGAAGGTAAACGGGTATATAGATGTGCG 3782  
Qy 788 -----ProSerProSerValSerProSerAlaSerProSerLeuSer 801  
Db 3783 ATAGTGTGGGAAGAGAGCCGAGCGGGGTACAAAGCCGCGGAGTAGTAAACACCGACA 3842  
Qy 802 ProSerProSerProSerSerProSerProSerProSerProSerProSerProSer 821  
Db 3843 CCGGACCGACCCCGACATCGACCGCGACACCAACCTTACCAACCACTGCACCGACA 3902  
Qy 822 SerSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 841  
Db 3903 TCAGCCCGCAGACCGGCCAACAGTGCAGCAACCGCGACTCCCAACCGCGCGAGCA 3962  
Qy 842 AlaSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 861  
Db 3963 GTGACGGTTACTGTGTACTCCGACACCGACACCA-----ACACCGAGC 4004  
Qy 862 ProThrProSerSerProValSerGly---GlyValLysValGlnTyrLysAsnAsn 880  
Db 4005 CCGACACCGACAGGAGACACCTGGCAGCGGGAAGTGGTTTGAAGGTTACTTACAAAGAACAT 4064  
Qy 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900  
Db 4065 GAGACAAGTGCAGCACCAAGTTCTATAAGCGCGTGGTTTAAGATAGTAGTAGGAGCAGC 4124  
Qy 901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 920  
Db 4125 AGCAGTGTGTATCTTAGCAGGTTTAAAGATAAGATACTGGTACACAGTGGATGGTACAAG 4184  
Qy 921 SerThrLeuValTyrAsnCysAspTrpAlaIleGlyCysGlyAsnIleArgAlaSer 940  
Db 4185 CCACAGAGTGGCGTA---TGTGACTGGGCACAGATAGGGGCAAGCAATGTGACTTCAAT 4241

QY 941 PheGlySerValIleProAlaThrProThrAlaAspThrTyrLeuGln 956  
 Db 4242 TTTGTGAAGCTGAGCAGCGAGTGAGTGAGCGGATTATTACTTTGGAG 4289

## RESULT 11

AAD26525  
 ID AAD26525 standard; DNA; 11707 BP.

XX AC AAD26525;

XX DT 26-MAR-2002 (first entry)

XX DE Active cellulase protein, celB gene.

XX KW Active cellulase protein; alkalophilic; textile processing; proteinase;  
 KW detergent additive; stonewashed appearance; cotton-containing denim;  
 KW CelB5; thermophilic; commercial detergent; celB gene; ds.

XX OS Unidentified.

XX FH Key Location/Qualifiers  
 FT CDS 6255..10535

FT /tag= a  
 FT /product= "CelB protein"

FT /tag= b

FT /product= "DNA encoding B4/5 protein"

FT /tag= c

FT /product= "DNA encoding B/5 protein"

XX US6294366-B1.

XX PN 25-SEP-2001.

XX 19-AUG-1998; 98US-00136574.

XX 19-SEP-1997; 97US-00932571.

XX (CLURN ) CLARIANT FINANCE BVI LTD.

XX Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;

XX Morgan H, Williams DP;

XX WPI; 2002-081780/11.

XX P-PSDE; AAE16323.

XX New cellulase active protein, useful in textile processing or commercial  
 PT detergents, e.g. for improving the feel or appearance of cotton-  
 PT containing fabrics, is stable under conditions of alkaline pH and  
 PT elevated temperatures.

XX Disclosure; Col 27-38; 61pp; English.

XX The present invention relates to a cellulase active protein, which is  
 CC substantially free of proteinases of native thermophilic and  
 CC alkalophilic origin, where the cellulase active protein consists of the  
 CC CelB5 amino acid sequence. The cellulase active protein is useful for  
 CC treating cellulosic materials including cotton-containing fabrics, as  
 CC detergent additives. The cellulase active protein is also useful for  
 CC improving the feel and/or appearance of cotton-containing fabrics, for  
 CC removing surface fibers from cotton-containing knits or for imparting  
 CC stonewashed appearance to cotton-containing denims. The present proteins  
 CC are stable under condition of alkaline pH and elevated temperatures, thus  
 CC suitable for textile processing and in commercial detergents. The present  
 CC sequence is celB gene

XX SQ Sequence 11707 BP; 3827 A; 1995 C; 2994 G; 2891 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 9,59e-06 Length: 11707  
 Score: 409,00 Matches: 263  
 Percent Similarity: 32.37% Conservative: 163

Best Local Similarity: 19.98% Mismatches: 386  
 Query Match: 7.96% Indels: 504  
 Db: 6 Gaps: 58

US-09-917-376-1 (1-957) x AAD26525 (1-11707)

QY 14 ArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGly 33  
 Db 735 AAGAGGGTAATTCAATTCTT---TCCTTATTGTTTTTTTAAATAACACGCTTGTAGGT 791  
 QY 34 ValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThrGlnProTyrThr 53  
 Db 792 ACTTTG-----ATATTCATCAGGAAGCAAAGCAGCAGCA-----TATACT 833  
 QY 54 TrpSerAenValAlaIleGlyGlyGlyPheValAspGlyIleValPheAenGluGly 73  
 Db 834 -----GTTGATTTGAAGGT 848

QY 74 AlaProGlyIleLeuTyr-----ValArgThrAspIleGlyGly 86  
 Db 849 GCTGATACATTTATCTTACTTTCTTATGGAATAATCGAGCATAGCAGTTGACATGGCAAT 908

QY 87 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 106  
 Db 908 -----

QY 107 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 126  
 Db 909 -----GCATATAATGGTAAAGTAGTGTCAGGGGTGTC-----AATAGAAGTTCA 953

QY 127 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146  
 Db 954 ATATGGGATGGAGTTGCAGTT-----CACGTTAA 983

QY 147 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThr-----158  
 Db 984 AACATTATGAACAATGGAAACCACTAGTGTTCAGCGTATGTAACCATAGTACCAG 1043

QY 159 ---ProLeuProPheLysLeu-----GlyGlyAsnMetProGlyArg 171  
 Db 1044 AAGCCGTTGTCATTGGTATCTCAGCGGTTTACGACGATGGAGTGGGTTAAGAGTACT 1103

QY 172 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 191  
 Db 1104 CTCATAGTGAGGTGTGGCTATT---CCAAAT-----1133

QY 192 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 211  
 Db 1134 -----TATTGGAAGAAATTT---1148

QY 212 AsnPheProAspValGlyTyrTyrIleAlaAsnProThrAspThrGlyTyrGlnSer 231  
 Db 1149 -----GTTGGTAAATGGACTCCAAATATTAGCAATGTCCAGGAATTTGTTAAT 1196

QY 232 AspileGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 251  
 Db 1197 GTAATACACACAAATGTAGAAAGCGAAGTAGATATAATGTTGAC-----1241

QY 252 SerLysThrIlePheValGlyValAlaAspProAen-----AsnProValPhe 267  
 Db 1242 -----TATATCCAATATATGATGATAATAGTTACCTATCAATGCAGTGACA 1289

QY 268 TrpSerArg-----AspGlyGlyAlaThr-----TrpGlnAla-----Val 279  
 Db 1290 TTTTCAAGTGGATTGTAAAGTGCGACTACCGAGGGTTGCGAGGCAAGGGAAGCGGTGT 1349

QY 280 ProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisVal 299  
 Db 1350 ACAGTAAACACAGATAGCGTTGTGGCATATATAGTGGCAATAGTAGT-----1394

QY 300 LeuTyrIleAla-----ThrSerAsnThrGlyGly-----ProTyrAspGly--- 313  
 Db 1395 TTGTACGTGAGTGAAGAACGTCATTAATTTGGCATGTTCCGCGTAGATACAAATT 1454

QY	314	---	Ser	Ser	Gly	Asp	Val	Trp	Lys	Phe	Ser	Val	-----	Thr	Ser	Gly	Thr	Trp	328
Db	1455	TTGGAA	CAG	GGT	AAAG	TGT	AT	AAAA	ATA	AGT	GT	TTGG	GT	TTAT	CAGA	ACA	CAGT	GT	1514
QY	329	Thr	Arg	Ile	Ser	Pro	Val	Pro	Ser	Thr	Asp	Thr	Ala	Asn	Asp	---	Tyr	Phe	347
Db	1515	CAAAA	AAAT	GT	CA	TAA	CTA	CTG	CA	AAAG	AAAT	TG	CT	CA	CA	AGT	CA	AA	1574
QY	348	Gly	Leu	Thr	Ile	Asp	Arg	Gln	His	Pro	Asn	Thr	Ile	Met	Val	Ala	Thr	Gln	367
Db	1575	AAT	CT	GAT	ATA	TAA	CAG	GAT	GT	ACC	GAT	ATA	AT	CGT	GGT	TG	AGT	CA	1634
QY	368	T	rp	Pro	Asp	Thr	Ile	ile	Phe	Arg	Ser	Thr	Asp	Gly	Gly	Ala	Thr	Trp	387
Db	1635	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1670
QY	388	Asp	Trp	Thr	Ser	Tyr	Pro	Leu	Arg	Ser	Leu	Arg	Tyr	-----	-----	-----	-----	-----	401
Db	1671	CTT	AT	TG	TTG	TG	AGC	CA	CA	AAAT	TG	CA	AAAT	TT	TG	CGT	TT	CGG	1730
QY	402	Asp	Ile	Ser	-----	Ala	Glu	Pro	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	407
Db	1731	GAT	TT	AT	CA	AG	TTG	CGT	G	A	CT	G	A	CT	G	A	CT	TT	1790
QY	407	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	407
Db	1791	GAT	TAT	T	TT	CA	A	AGT	AGT	AGT	AGT	AGT	AGT	AGT	AGT	AGT	AGT	AGT	1850
QY	407	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	407
Db	1851	AAG	AT	GGT	TTT	TG	A	AGC	ATT	T	CA	A	T	AGT	AT	T	AGT	AT	1910
QY	408	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	412
Db	1911	TTA	CTT	GT	CGAT	GAA	AAT	ACT	T	CA	ACT	TT	T	AG	CA	A	AGC	AGC	1970
QY	413	Val	Gln	Pro	Asn	Pro	Pro	Val	Pro	Ser	Pro	Leu	Ser	Gly	Trp	Met	Asp	Glu	430
Db	1971	ACA	AGT	TAA	CA	CA	ATT	TG	CCAT	CAG	AGT	CTA	CAC	ACT	TG	TTT	GGC	AT	2030
QY	430	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	430
Db	2031	TG	GTTTT	TTT	CA	AG	GAC	AC	AAAT	TG	GAAT	AT	AC	GT	TG	CA	AG	AGT	2090
QY	430	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	430
Db	2091	AA	CAG	AT	TAT	T	T	AT	T	AC	G	T	AG	T	AG	G	A	G	

Qy	526	TyrAlaGluLeuAsnProSerIlelleValArgAlaGlySerPheAspProSerSerGln	545
Db	2475	ATTACGGAGCTTGCATATGAGTTTATTCAGTGGGGTTTCGAGTACCAGTTACTCAACGCCA	2534
Qy	546	ProAsnAsp	548
Db	2535	CCAGAGATCTCTGTATAAAACAGGCCGATCAGATATAAGGAGTTATTTGATTTGTTTAA	2594
Qy	549	-----ArgHisValAlaPhe	553
Db	2595	AAGTACAACAATGTAATAAACAGGTGAACACTCTGGGGACTGAAGGATGATTACTCATGG	2654
Qy	554	---SerThrAspGlyGlyLeuAsn	560
Db	2655	CTGAGTCAAAACACTTTGGAAAAAGGTATTACCCGTTGTATTATTTGATGAAAACTATAATCA	2714
Qy	561	-----TrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAla	577
Db	2715	AAATATGCCCTTTTGGAGCCTGATTGAGCCAACTGTGTATACGGGCCAACTCAACANTGCCA	2774
Qy	578	Ala	586
Db	2775	GCACCACCGCTATTCAAATACCTACACCAACTCCACACCAACCCCGACACCGACAGTG	2834
Qy	587	TrpAlaProGlyAspProGlyGlnProValValTyAlaValGlyPheGlyAsnSerTrp	606
Db	2835	AGTGCAACGCCAACACCAAGCA---CCGACGGCATCACCGGTAGGT---GGCAGTTACTGG	2888
Qy	607	AlaAlaSerGlnGlyValProAla	618
Db	2889	ACGCCAGTGAGAGTTACAGTGGCGTGAAAGGTATGTTATGCGAATGGGAATTTAAGCAGC	2948
Qy	619	ArgSerAspArgValAsnProGlyThrPheTyAlaLeuSerAsn	635
Db	2949	CCGACGAATGTATTGAATCTTAG-----ATAAGATAGAGAAATGTTGGACGACACGC	3002
Qy	636	-----PheTyArgSerThrAspGlyGlyValThr	645
Db	3003	GTAGATCTTAGCAGGGTGAAGTAAGTACTGTCACCGATAGTGTGGAGCAACACAG	3062
Qy	646	-----PheGlnPro	651
Db	3063	AGTGTAAGTGTAAACAGCAGCATAGATCTCTCGCTATATAGATGTGAAGTTGTGAAGCTT	3122
Qy	652	GlyLeuProSerSerGlyAla	660
Db	3123	GGACGAAACGACGGCGAGCGGATTACTATGTGGAGTAGGCTTTAAGAGTGGNCGAGGG	3182
Qy	661	ValMetPheHisValProGlyLysGluGlyAspLeuTrpLeuAlaSerSerGly	680
Db	3183	GTTTTGGCAGCAGCGGCAACAGCAGAGGAGATAAGACTTAGCATACAGAAAGGCGAGTGGC	3242
Qy	681	LeuTyRHisSerThrAsnGlyGlySerSerTrpSerAla	693
Db	3243	AGCTACAATCAGTCAAAATGACTATTCGGTGAGAGTGCACACAGCTATATAGAGAACGAG	3302
Qy	694	--IleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySer	712
Db	3303	AAGTAAACGGGTATATAGATGATGTACTTGTATGGGAAGACAGACGCGACGAGAACGCC	3362
Qy	713	SerTyRProAlaValPheValValGlyThrIleGlyValThrGlyAla	729
Db	3363	CAGATCAAGGTATGGTATGCAATGGGAATTTAAGCAGCCCGAATGATTGGAATCCT	3422
Qy	730	---TyrArgSerAspAspCysGlyThrThr	738
Db	3423	AAGATAAAGATAGAGAATGTTGGGACGACACAGCGGTAGATCTTAGCAGGGGTGAAGGTAA	3482
Qy	739	---TrpValLeuIleAsnAspAspGlnHisGlyTyrGlyAsnTrpGlyGlnAlaIleThr	757
Db	3483	TACTGGGTACAGTATAGTGTGGGCAACACAGAGTGAAGTGTAAACAGCAGCAGTAAC	3542
Qy	758	GlyAspHisAlaAsnLeuAArgValTyIleGlyThrAsnGlyArgGlyIle	776

```
Db 3543 CTTGGCTATATAGATGTAAGATTGTTGTAAGCTTGGACCAATGAGTGGCGGATTAC 3602
Qy TyrGlyAspIleGly-----GlyVala----- 783
Db 3603 TATGTGGAGATAGCTTTAAGATGTGGAGCAGGGGTTTGGCAGCAGGGCAGACGCAAG 3662
Qy 784 -----ProSerGlySer----- 787
Db 3663 GAGATAAGACTTAGCATACAGAAGGGCAGTGGCAGCTACAATCAGTCAAATGACTATTGC 3722
Qy 787 ----- 787
Db 3723 GTGAGGAGTCAACAGCTATATAGAACGAGAGGTAACGGGGTATATAGATGGCG 3782
Qy 788 -----ProSerProSerValSerProSerAlaSerProSerLeuSer 801
Db 3783 ATAGTGTGGGAGAGAGCCGAGCGAGGGGTACAAAGCCGGGGAGTAGTAACACCGACA 3842
Qy 802 ProSerProSerProSerSerProSerProSerProSerProSerProSer 821
Db 3843 CCGGACCGACCCGACATCGACCGGACACCAACCTACAACACCTGCGACCGACA 3902
Qy 822 SerSerProSerProSerProSerProSerProSerProSerProSerProSer 841
Db 3903 TCAGCCCGACACCGAGCCCAACAGTACAGCAACGCCGACTCCACACCGCGCGCGACA 3962
Qy 842 AlaSerProSerProSerProSerProSerProSerProSerProSerProSer 861
Db 3963 GTACGGTACTGTGACTCGACACCGACACCA-----ACACCGACG 4004
Qy 862 ProThrProSerSerProValSerGly---GlyValLysValGlnTyrLysAsnAsn 880
Db 4005 CCACACCGACAGGACACTCGCACCGGGAAGTGGTTGAAGTACTATACAGAACAAT 4064
Qy 881 AspSerAlaProGlyAspAsnGlnLysProGlyLeuGlnValValAsnThrGlySer 900
Db 4065 GAGACAAAGTCGAGCACAAGTCTTATAAGCCCGTGGTTAAGTAGTAGTGAATGAGGCGCAG 4124
Qy 901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 920
Db 4125 AGCAGTGTGATCTTAGCAGGGTTAAGATAAGATATCTGTGTACACAGTGGATGGTGAACAG 4184
Qy 921 SerThrLeuValTyrAsnCysAspTTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSer 940
Db 4185 CCACAGAGTCGGTA---TGTGACTGGGCACAGATAGGGGCAAGCAATGTGACATCAAT 4241
Qy 941 PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 956
Db 4242 TTTGTGAGCTGAGCAGCGAGTGGAGTGGAGCGGATATTACTTTGGAG 4289
RESULT 12
AAF14988
ID AAF14988 standard; cDNA; 1103 BP.
XX
AC AAF14988;
XX
XX 15-SEP-2003 (revised)
DT 13-MAR-2001 (first entry)
XX
XX Trichoderma reesei EST SEQ ID NO:7511.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Hypocrea jecorina.
XX
XX WO200056762-A2.
XX
XX 28-SEP-2000.
```

```
XX 22-MAR-2000; 2000WO-US007781.
XX
XX 22-MAR-1999; 99US-00273623.
XX (NOVO ) NOVO NORDISK BIOTECH INC.
XX (NOVO ) NOVO NORDISK AS.
XX
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX WPI; 2000-594572/56.
XX
XX Claim 89; Page 3034; 3161pp; English.
XX
XX The present invention describes a method for monitoring differential
XX expression of genes in a first filamentous fungal (FF) cell relative to
XX expression of the same genes in one or more second filamentous fungal
XX cells. The method uses fluorescence-labeled nucleic acids isolated from
XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX are used in the methods for monitoring differential expression of genes
XX in a first filamentous fungal (FF) cell relative to expression of the
XX same genes in one or more second filamentous fungal cells. Monitoring the
XX global expression of genes from FF cells allows the production potential
XX of the microorganisms to be improved. New genes may be discovered,
XX possible functions of unknown open reading frames can be identified and
XX gene copy number variation and stability can be monitored. The expression
XX of genes can be used to study how FF cells adapt to changes in culture
XX conditions, environmental stress, spore morphogenesis, recombination,
XX metabolic or catabolic pathway engineering. Using ESTs provides several
XX advantages over genomic or random cDNA clones including elimination of
XX redundancy as one spot on an array equals one gene or open reading frame,
XX CC and organisation of the microarrays based on function of the gene
XX products to facilitate analysis of the results. AAF07478 to AAF11247
XX represents ESTs from Fusarium venenatum; AAF11248 to AAF11853 represents
XX ESTs from Aspergillus niger; AAF11854 to AAF14878 represents ESTs from
XX Aspergillus oryzae; and AAF14879 to AAF15337 represents ESTs from
XX Trichoderma reesei, which are all specifically claimed in the present
XX invention. (Updated on 15-SEP-2003 to standardise OS field)
XX
SQ Sequence 1103 BP; 230 A; 322 C; 307 G; 223 T; 0 U; 21 Other;

Alignment Scores:
Pred. NO.: 5.08e-06 Length: 1103
Score: 392.50 Matches: 114
Percent Similarity: 45.05% Conservative: 36
Best Local Similarity: 34.23% Mismatches: 125
Query Match: 7.64% Indels: 58
DB: 3 Gaps: 10

US-09-917-376-1 (1-957) x AAF14988 (1-1103)
Qy 348 GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp 367
Db 3 GGCCTTGGCTCGATTGTCGAAAGCCAGGACCCCTTGTGTGTCTTTGAACCTCTGG 62
Qy 368 TrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp 387
Db 63 TGGCCAGATGCTCAGCTGTGTTGCGTGCACCGACTCTGGGACACATGGAGCCGATCTGG 122
Qy 388 AspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluPro 407
Db 123 GCGTGGCGAGCTATCCGACTGAGACCTTATTACTACAGCATCTCAACTCCCAAGCACCG 182
Qy 408 TrpLeuThrPheGly---ValGlnProAsnProProValProSer-----Pro 422
Db 183 TGGATCAAGAACAACTTTATCGATGTGTGACGCGAGTCCACCGTCGCGATGGTCTNATCAAG 242
Qy 423 LysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeu 442
XX
```

243	Db	CGCCTCGGCTGGATGATGTAGTCTNTCGAGATTGACCCACCCGACGCAANNACTGGCTT	302
443	QY	-TyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer-	460
303	Db	TTACGGCACCGGATGACAAATNTTGGCGGCCAGATTCAACAACTGGGACACGCGCC	362
461	QY	-----GlyGlyGlnIleHisIleAlaProMetValLys-GlyLeuGluGluThrAlaVal	478
363	Db	ACAATGTGGTCAATCCAAATTACTTGGCAGACGGGATTCGAAGGAATT-----TTTCGCT	416
478	QY	alaAspLeuIleSerProProSerGly-----	487
417	Db	TCAAGGACCTGGCCCTTTTTCACCCGGGGGAAAGCAGAGCTTTTGGCGCGCAAGTCGGGAGACG	476
488	QY	-----AlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspV	505
477	Db	ANCAACGGGTTTACCTTTTGGCCAGCAGAAACGACCTTGGG-----	517
505	QY	alThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThr-SerVal	524
518	Db	-----ACATTGCCCGCACAGCGTTTGGCGNACTCCACATGGGCGCACCTCGACGAAGCGTC	572
525	QY	AspTyrAla--GluLeuAsnProSerIleIleValArgAlaGlySer-----	539
573	Db	GACTACGCGGGAACCTCGTCAAGAGCGCTTCGTCCGGTCGGCAACACCGCGCGGCACG	632
540	QY	-----PheAspProSerSerGlnProAsnAspArgHisValAlaPheS	554
633	Db	CAACAAGTGGCCATTTTTTTCOGAAGCGGGCGCGACGTNGGAA-----GCAATTTC	686
554	QY	erThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyG	574
687	Db	GAACCTAACGCT-----GGNTCCGAACACCGTTTCCATTCAAAACGGCG	728
574	QY	lyThrValAlaIleSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyG	594
729	Db	GCOCGGTGGCCTATTCCGCGCAGCGCGCACAGATCTCTGTGTCGACCGCCTCGTCCGCG-	787
594	QY	lnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProA	614
788	Db	-----GTGACGCGCTCGCAGTTTCAGGGCAGCTTTGGCTCTCGCTCTCGAGCCTGCCCCG	839
614	QY	laAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnG	634
840	Db	CGGGCGCGCTCATCGNCTCGGACGAAGAGCAACAGCGGTNTTCTACGCGCGCTCCGGAT	899
634	QY	lyThrPheTyrArgSerThrAspGlyGly	643
900	Db	CGACCTTTTACGTCAGCAAGGACACCGCG	928

RESULT 13  
AA894956/c  
ID AA894956 standard; cDNA; 651 BP.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

P-PSDB; ABG20769.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

Claim 1; SEQ ID NO 20760; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of states expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AA564197-AA594564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at <http://wipo.int/pub/published> pct sequences

Sequence 651 BP; 152 A; 86 C; 241 G; 172 T; 0 U; 0 Other;

**Alignment Scores:**

Pred. No.:	2.92e-05	Length:	651
Score:	365.50	Matches:	101
Percent Similarity:	55.15%	Conservative:	8
Best Local Similarity:	48.33%	Mismatches:	44
Query Match:	7.12%	Indels:	57
DB:	5	Gaps:	8

US-09-917-376-1 (1-957) x AAS84956 (1-651)

671	GlyAspLeuTrp---	LeuAlaAlaSerSerGly	-----	LeuTyrHisSerThrAsn	686
	:::	:::	:::	:::	
538	GGAAAGCTCTGGGTCTCGCTGAAGCTTCAGGCCGCGTATGCGCAATCATGTACGTGA	479			
687	GlyCysSerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly	706			
	:::	:::	:::	:::	
478	GCAGGGCGGTAGCTGGCGATGT	-----	---	TGA	455
707	LysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyVal	726			
	:::	:::	:::	:::	
454	AAGCACCACCGAGCC	-----	CCATGCTCCCGATCTCTGGTACAGCTGGC	-----	410
727	ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGln	746			
	:::	:::	:::	:::	
409	-----AGGACAGCTCACTGTC-ACCACA	-----	-----	-----	387
747	HisGlnTyrGlyAsnTrp	-----	GlyGlnAlaIleThrGly-AspHi	760	
	:::	:::	:::	:::	
386	-----TAGAACTGGCAGAGGGCATGGCATGGAGCGGCCATCAGAGGAAGATCA	336			
760	sAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle	780			
	:::	:::	:::	:::	
335	TCTCCCATCACCATAACCGTCGCATCTCCACA	-----	-----	-----	305
	:::	:::	:::	:::	







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 14, 2004, 08:53:05 ; Search time 5623 Seconds  
(without alignments)  
5082.359 Million cell updates/sec

Title: US-09-917-376-1  
Perfect score: 5135  
Sequence: 1 MDRSENIRLTMRRSLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=xlh  
-Q=/cgn2/1/USPTO.spool/US09917376/runat\_14052004\_085257\_4630/app\_query.fasta\_1.1095  
-DB=EST -QPMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09917376 @CGN\_1\_1\_4087@runat\_14052004\_085257\_4630 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query %	Length	DB	ID	Description
1	631.5	12.3	704	14	CF872104	trico29xo
2	631.5	12.3	761	14	CB902332	trico29xo
3	603.5	11.8	747	14	CF876916	trico74xd
4	603.5	11.8	814	14	CB905388	trico74xd
5	415.5	8.1	707	14	CF880713	trico82xn
6	415.5	8.1	782	14	CB907625	trico82xn
7	401.5	7.8	693	14	CF882065	trico29xo
8	362	7.0	748	29	CE342004	tigr-gss-
9	361.5	7.0	751	14	CF868882	trico16xm
10	361.5	7.0	803	14	CB898982	trico16xm
11	360	7.0	791	29	CNS0280T	Tetraodon
12	355	6.9	522	29	CE190185	tigr-gss-
13	353	6.9	493	28	CC066065	CSU-K33r.
14	350	6.8	694	28	CC119324	NDL_70K21
15	347	6.8	259	28	AZ374273	LM0126G21
16	347	6.8	587	13	C79507	C79507 Mous
17	338	6.6	738	14	CF866388	trico06xe
18	338	6.6	794	14	CB896354	trico06xe
19	333	6.5	703	29	CE176982	tigr-gss-
20	329.5	6.4	476	13	C79500	Mous
21	328	6.4	562	28	AZ855343	2M0159A22
22	314.5	6.1	546	10	BF072664	NCSM3H73
23	314	6.1	619	28	BZ192350	CH230-248
24	314	6.1	959	29	CNS022KT	Tetraodon
25	310	6.0	744	28	BZ277113	CH230-450
26	307	6.0	538	29	CC847893	NDL_4416
27	306.5	6.0	389	13	BUG39045	mgcwo11xd
28	306	6.0	961	29	CNS03QKT	AL256070 Tetraodon
29	301.5	5.9	464	28	AZ928926	479.dif19
30	298.5	5.8	440	28	AZ312209	1M0027L19
31	297	5.8	929	14	CF885920	trico85xf
32	292	5.7	393	28	AZ332252	1M0060H12
33	291	5.7	761	29	CNS041RC	AL270561 Tetraodon
34	287	5.6	559	28	BH306809	CH230-100
35	286.5	5.6	713	14	CF875815	trico39xx
36	286.5	5.6	782	14	CB904767	CB904767 trico39xx
37	285	5.6	534	28	AZ014305	RPC1-23-3
38	284	5.5	592	28	AZ015337	RPC1-23-3
39	284	5.5	632	29	CE696112	tigr-gss-
40	282.5	5.5	719	14	CF881775	trico85xf
41	282.5	5.5	775	14	CB908435	trico85xf
42	282	5.5	647	9	AU061620	AU061620
43	274	5.3	401	12	BG887959	EST513810
44	272.5	5.3	697	28	AZ374916	1M0128I06
45	270	5.3	400	28	AQ171567	HS_3163_B

# ALIGNMENTS

RESULT 1  
CF872104  
LOCUS  
DEFINITION trico29xo13.b1 T.reesei mycelial culture, Version 6 October 2003  
ACCESSION CF872104  
VERSION CF872104.1 GI:38126786  
KEYWORDS EST.  
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)  
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Hypocrea; Hypocrea.  
REFERENCE 1 (bases 1 to 704)



```
Db 146 CATGCTGCC-----TTTTATGGAAGACGTCAAGCTCGCGCGCGCGGCGG 193
Qy 64 PheValAspGlyLeuValPheAsnGluGlyAlaProGlyLeuTyrValArgThrAsp 83
Db 194 TTTCGTCCTCCCGGATCATCTTCCATCCCAAGACAAAGGCGTAGCATATGCAAGACAGAT 253
Qy 84 IleGlyGlyMetTyrArg-----TrpAspAlaAlaAsnGlyArgTrpLeuPro-----Leu 100
Db 254 ATTGGCGGGGTGTACCGCTCTTCTCCGACACTCATGGACCGCC-GTCAAGGATGGGATT 312
Qy 101 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp 120
Db 313 GCTGATAAATGCCGCTGGGCAAACTGG-----GGCATCGACGCTGTGGCTTTGAT 363
Qy 121 ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPro 140
Db 364 CCGCAGGACGATCAAAAGGTGTATGCCGAGTCGGCATGTATACGAACAGCTGGGATCCG 423
Qy 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu 160
Db 424 AGTAATGGAGCCATCATCTCGTCGTAGACCGCGGCGCAACGTGGTCTTCCACCAACTTG 483
Qy 161 ProPheIleLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180
Db 484 CCCTTCAAAATCGGGGGTAAATGATCCAGGACGCGGAGCGCGAGAGCGTCTGGCTGTGAT 543
Qy 181 ProAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpAspSer 200
Db 544 CCGGCCACTCCAAATCATCTACTTTGGTGTCTGCTCAGGAACGGCTCTCGAAGTCT 603
Qy 201 ThrAspSerGlyAlaThrTrpSerGlnMetMetAsnPheProAspValGlyThrTrpIle 220
Db 604 ACGGACGGCGGCGTGACCTTTCCAAAGTCTCGTCTGTTACGGCAACTGGGAGTATCATC 663
Qy 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla 240
Db 664 CCAGACCCGAGTATTCACACGGCTACCAAGCGATCAGCAGGATCAGCAGGACTCATGTGGTTACG 723
Qy 241 PheAspLysSerSerSerSerLeuGlyGlnAlaSer 252
Db 724 TTGACTC-ACCAGCAGCAGCAGCGGNGGAGGCCACG 758
```

```
RESULT 3
CF876916 747 bp mRNA linear EST 31-OCT-2003
LOCUS trico074xd02.b1 T.reesei mycelial culture, version 6 October 2003
DEFINITION Hypocrea jecorina cDNA clone trico074xd02, mRNA sequence.
ACCESSION CF876916
VERSION CF876916.1 GI:38131598
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE 1 (bases 1 to 747)
AUTHORS Diener S.E., Dankmeyer L., Dunn-Coleman N., Houfek T.D.,
Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and
Dean, R.A.
TITLE Analysis of the protein processing and secretion pathways in a
Trichoderma reesei EST dataset
JOURNAL Unpublished (2003)
COMMENT Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu
Seq primer: Tr-F1 primer.
Location/Qualifiers
1..747
/organism="Hypocrea jecorina"
/mol_type="mRNA"
```

```
/strain="QM6a"
/db_xref="taxon:51453"
/clones="trico074xd02"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October 2003"
/note="Vector: pREP3Y; Site_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."
```

## ORIGIN

```
Alignment Scores: 4.43e-13 Length: 747
Pred. No.: 603.50 Matches: 121
Score: 65.43% Conservative: 38
Percent Similarity: 49.79% Mismatches: 64
Best Local Similarity: 11.75% Indels: 20
Query Match: 14 Gaps: 6
DB:

US-09-917-376-1 (1-957) x CF876916 (1-747)

Qy 25 SerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
Db 61 TCTGAGTCTTGCCTTGTCTTGGGCGCCGTC-----ATCCTGCTCC 102
Qy 45 HisAlaAlaThrThrGlnProTyrTrpSerAsnValAlaIle---GlyGlyGlyGly 63
Db 103 CATGCTGCC-----TTTTCATGGAAGAACGTCAAGCTCGCGCGCGCGCGGCG 150
Qy 64 PheValAspGlyLeuValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp 83
Db 151 TTGTCCTCCCGCATCATCTTCCATCCCAAGACAAAGCGTAGCATATGACGAACAGAT 210
Qy 84 IleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
Db 211 ATTGGCGGGTGTACCGCTCAAC---GCCAGCACTCATGGACCGCGCTCAGGATGGG 267
Qy 104 Val-----GlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAla 119
Db 268 ATTGCTGATATGCGGCTGGCAACTGG-----GGCATCGAGCTGTTCGGCTT 318
Qy 120 AspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp 139
Db 319 GATCCGCGAGCAGCATCAAAAGGTGTATGCGCGAGTCGCGCATGTATACGAACAGCTGGAT 378
Qy 140 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 159
Db 379 CCGAGTAATGGAGCCATCATCTCGCTCGTCAGACCGCGCGCAACGTGTCTTCACCAAC 438
Qy 160 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 179
Db 439 TTGCGCTTCACAGTCGGGGGTAAATGTCAGAGACCCGAGCGCGGAGAGCGTCTGGCTGC 498
Qy 180 AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArg 199
Db 499 GATCCGGCCACCTCCAAACATCATCTACTTTGGTGTCTCGCTCAGCAGACGGCTCTGGAAG 558
Qy 200 SerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyr 219
Db 559 TCTACCGACGCGCGGTGACCTTTTCCAAGGCGCTCGTCGCTCAGCGCCCTGGGACGCTCC 618
Qy 220 IleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpVal 239
Db 619 CTCCTCCAGCCCGAGTGTATCCACGGCTACCAACCGCGACAGCAAGCGGCTCATGTGGGT 678
Qy 240 AlaPheAspLysSerSerSerSerLeuGly-GlnAlaSerLysThrIlePheValGlyVa 259
Db 679 ACGTTGACTCACCAGCATCACACCGGTGGAGCCAGCGCTCGTATTTTGTCTGCCAC 738
Qy 259 lAlaAsp 261
Db 739 GACTGAC 745
```

```

RESULT 4
CB905388      814 bp  mRNA  linear  EST 02-JUL-2003
LOCUS        trico074xd02 T.reesei mycelial culture, Version 3 april Hypocrea
DEFINITION   jecorina cDNA clone trico074xd02, mRNA sequence.
ACCESSION    CB905388
VERSION      CB905388.1 GI:30120046
KEYWORDS     EST.
SOURCE       Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM     Hypocrea jecorina
              Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
              Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE    1 (bases 1 to 814)
AUTHORS      Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
              Dunn-Coleman,N.S., Goedegebuur,E., Houfek,T.D., England,G.J.,
              Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,
              Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.
              Transcriptional regulation of biomass-degrading enzymes in the
              filamentous fungus Trichoderma reesei
              J. Biol. Chem. 278 (34), 31988-31997 (2003)
JOURNAL      22803314
MEDLINE      12788920
PUBMED
COMMENT      Contact: Pamela K. Foreman
              Genencor Intl.
              925 Page Mill Road, Palo Alto, CA 94304, USA
              Tel: (650) 846-7635
              Fax: (650) 621-7817
              Email: pforeman@genencor.com
              Seq primer: LT-P1 primer:
              Location/Qualifiers
                1..814
                /organism="Hypocrea jecorina"
                /mol_type="mRNA"
                /strain="QM6a"
                /db_xref="taxon:51453"
                /clone="trico074xd02"
                /dev_stage="mycelia"
                /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
                culture grown from 24 hrs to 6 days with varying Carbon
                and Nitrogen sources and concentrations."

ALIGNMENT SCORES:
Pred. No.: 4,85e-13 Length: 814
Score: 603.50 Matches: 121
Percent Similarity: 65.43% Conservative: 38
Best Local Similarity: 49.79% Mismatches: 64
Query Match: 11.75% Indels: 20
DB: 14 Gaps: 6

US-09-917-376-1 (1-957) x CB905388 (1-814)

QY 25 SerPheAlaValAlaAlaLeuGlyValLeuProIleAlaThrAlaSerProAla 44
Db 128 TCTCGAGTCCTTGCCTTGTCTGGGGCGGTC-----ATCCCTGCC 169
QY 45 HisAlaAlaThrGlnProTyrThrTTPSerAsnValAlaLe-----GlyGlyGly 63
Db 170 CARGCTGCC-----TTTTCATGAAGAAGACGTCAAGTCGCGCGCGCGCGC 217
QY 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp 83
Db 218 TTGCTCCCGGATCATCTTCATCCCAAGAAAGGCGTAGCATATGACGACAGAT 277
QY 84 IleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
Db 278 ATTGGCGGCTGTACCGCCTCAAC---GCCGACGACTCATGGACCGCGCTCAGCATGGG 334
QY 104 Val-----GlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAla 119
Db 335 ATTGCTGATAATGCCGCTGGCAACTGG-----GCCATCGAGCTGTTCGCTT 385

```

```

QY 120 AspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp 139
Db 386 GATCCGCGAGCATGATCAAAAGGTGTATGCGCATGTCGGCATGTATACGAACAGCTGGGAT 445
QY 140 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 159
Db 446 CCGAGTAATGGAGCCATCATTCGTCGTACAGCCGCGGCGCAACGTGGTCTTCACCAAC 505
QY 160 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 179
Db 506 TTGCGCTTCACAGTCGGGGTTAATGTCAGGACCCGCGAGCGGAGAGCGTCTGGCTGTC 565
QY 180 AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArg 199
Db 566 GATCCGCGCCACCTCAACATCATCTACTTTGGTGTCTGCTCACGACACGCGCTCTGGAAG 625
QY 200 SerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyr 219
Db 626 TCTACCGACGCGCGCGTACCTTTCCAGGCGCTGCTGCTCACGCGCCCTGGGAGCTCC 585
QY 220 IleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpVal 239
Db 686 CTCACGACCGGAGTGATTCACACGGCTACCAACGCGGACAAAGCAAGCGCTCATGTGGTT 745
QY 240 AlaPheAspLysSerSerSerSerLeuGly-GlnAlaSerLysThrIlePheValGlyVa 259
Db 746 ACGTTGACTCACCCAGCATCACGACGCGTGGCAGCCACGCGCTCGTATTTTGTCTGCCAC 805
QY 259 LAlaAsp 261
Db 806 GACTGAC 812

RESULT 5
CF880713      707 bp  mRNA  linear  EST 31-OCT-2003
LOCUS        trico082xnl9.b1 T.reesei mycelial culture, Version 6 October 2003
DEFINITION   Hypocrea jecorina cDNA clone trico082xnl9, mRNA sequence.
ACCESSION    CF880713
VERSION      CF880713.1 GI:38135395
KEYWORDS     EST.
SOURCE       Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM     Hypocrea jecorina
              Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
              Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE    1 (bases 1 to 707)
AUTHORS      Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D.,
              Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and
              Dean,R.A.
              Analysis of the protein processing and secretion pathways in a
              Trichoderma reesei EST dataset
              Unpublished (2003)
              Contact: Ralph A. Dean
              Fungal Genomics Laboratory
              North Carolina State University
              Campus Box 7251, Raleigh, NC 27695, USA
              Tel: 919-513-0020
              Fax: 919-513-0024
              Email: ralph.dean@ncsu.edu
              Seq primer: LT-F1 primer.
              Location/Qualifiers
                1..707
                /organism="Hypocrea jecorina"
                /mol_type="mRNA"
                /strain="QM6a"
                /db_xref="taxon:51453"
                /clone="trico082xnl9"
                /dev_stage="mycelia"
                /clone_lib="T.reesei mycelial culture, Version 6 October
                2003"
                /note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
                culture grown from 24 hrs to 6 days with varying Carbon
                and Nitrogen sources and concentrations."

ORIGIN

```

Alignment Scores:		
Pred. No.:	2,668-06	Length: 707
Score:	415.50	Matches: 92
Percent Similarity:	55.51%	Conservative: 34
Best Local Similarity:	40.53%	Mismatches: 90
Query Match:	8.09%	Indels: 11
DB:	14	Gaps: 3

US-09-0917-376-1 (1-957) x CF880713 (1-707)

423	Qy	LysLeuGlyTrpMetAspGluAlaMetAlaIleAspPropheAsnSerAspArgMetLeu	442
		:::	
33	Db	CGCCTCGCGTGGATGTGATCTTCCTCGAGATGTGACCAACCGCAGCAACCATGGCTC	92
		:::	
443	Qy	TyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGly	462
		:::	
93	Db	TACGGCACCGGAATGACAATCTTTGGCGGCACGATCTCACCAACTGGGACACCGCCAC	152
		:::	
463	Qy	GlnIleHisIleAlaProMetValIlysglyLeuGluGluThrAlaValAsnAspLeulle	482
		:::	
153	Db	AATGTGTCAATCCAATCACTGGCAGACGGCATCGAGGAATCTTCGTCCAGCACTGGCC	212
		:::	
483	Qy	SerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis	502
		:::	
213	Db	TCTGCACCCGCGGAAGCAGCTATTGGCCGAGTCGAGACGACACACGGCTTACCTTT	272
		:::	
503	Qy	Ala-----AspValThrAlaValProSerThrIlePheThrSerProValPheThr	519
		:::	
273	Db	GCCAGCAGAAACGACCTTCGGGACATCGCCGAGACGGTCTGGGCAACGCCACATGGGCC	332
		:::	
520	Qy	ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySer	539
		:::	
333	Db	ACCTCGACGAGCGTCGACTACGCCGGAGACTCGGTCAAGAGCGTCTCGCGTCGGCAAC	392
		:::	
540	Qy	PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys	559
		:::	
393	Db	ACCCCGGCACGCACAG-----GTGGCCATCTCTGTCGACGGCGCGCGCG	437
		:::	
560	Qy	AsnTrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAlaSer	579
		:::	
438	Db	ACGTGGAGCATCGACTACCGCGGCCGACACGTCCTAAGAACCGCGGACGGTGGCTATTCTG	497
		:::	
580	Qy	AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla	599
		:::	
498	Db	GCCACGGCGCACACGATCCCTCTGTGTGCACCGCCTCTGTCGGC-----GTGCAGCGC	548
		:::	
600	Qy	ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg	619
		:::	
549	Db	TCGCAGTTCAGGGCAGCTTTGCTCTCGTCTGAGCCTTGCCCGGGGGCGCGTCACTGCC	608
		:::	
620	Qy	SerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSer	639
		:::	
609	Db	TCGCACAAGAAGACCAACAGCGTCTCTACGCCGCGCTCCGGATCGACCTTTTACGTACG	668
		:::	
640	Qy	ThrAspGlyGlyValThrPhe	646
		:::	
669	Db	AAGGACACCGCGCAGCAGCTTC	689
		:::	

RESULT 6	CB907625	782 bp	linear	EST 02-JUL-2003
LOCUS	CB907625			
DEFINITION	trico082xn19 T.reesei mycelial culture, Version 3 april Hypocrea jecorina cDNA clone trico082xn19, mRNA sequence.			
ACCESSION	CB907625			
VERSION	CB907625.1	GI:30122283		
KEYWORDS	EST.			
SOURCE	Hypocrea jecorina (anamorph: Trichoderma reesei)			
ORGANISM	Hypocrea jecorina			
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.			
REFERENCE	1 (bases 1 to 782)			
AUTHORS	Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,			

Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J., Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C., Oliveras, H.A., Teunissen, P., Yao, J. and Ward, M.  
Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*  
J. Biol. Chem. 278 (34), 31988-31997 (2003)  
22803314  
12788920  
Contact: Pamela K. Foreman  
Genencor Intl.  
925 Page Mill Road, Palo Alto, CA 94304, USA  
Tel: (650) 846-7635  
Fax: (650) 621-7817  
Email: pforeman@genencor.com  
Seq primer: LT-F1 primer.

**FEATURES**  
**source**

```

location/Qualifiers
1. .782
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="tric082xn19"
/dev_stage="mycelia"

```

## ORIGIN

Alignment Scores:		
Pred. No.:	2.95e-06	782
Score:	415.50	Matches:
Percent Similarity:	55.51%	Conservative:
Best Local Similarity:	40.53%	Mismatches:
Query Match:	8.09%	Indels:
DB:	14	Gaps:

US-09-917-376-1 (1-957) x CB907625 (1-782)

423	Qy	LysLeuGlyTrpMetAspGluAlaMetAlaIleAspPropheAsnSerAspArgMetLeu	442	Db	CGCTCGGGTGGATGATTGAGTCTTCGAGATTGACCAACCGCAGCAACACCTGGCTC	167
443	Qy	TyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGly	462	Db	TACGGCACCGGAATGACAATCTTTGGCGGCCACCATCTCCACCACTGGGACACGGCCAC	227
463	Qy	GlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIle	482	Db	AATGTGTCAATCCAATCACTGGCAGACGGCATCGAGGAATCTTCGTCCAGACCTGGCC	287
483	Qy	SerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis	502	Db	TCTGCACCCCGCGGAGCGAGCTATTTCGCCGACGTCGGAGACGACCAACGGCTTCACCTTT	347
503	Qy	Ala-----AspValThrAlaValProSerThrIlePheThrSerProValPheThr	519	Db	GCACGACAGAAACACCTTCGGGACATCCGCCGACAGCGTCTGGGCACGCCCACTGGGGCC	407
520	Qy	ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySer	539	Db	ACCTCGACGAGCGTGCAGTACGCCGGGAACCTCGGTCAAGAGCGTCTCGCGCTCGGCAAC	467
540	Qy	PheAspProSerSerGlnProAsnAspArgHisValIaPheSerThrAspGlyGlyLys	559	Db	ACGCGCGGCACCAACAG-----GTGGCCATCTCTGTCGACGCGCGCGCGG	512
560	Qy	AsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer	579	Db	ACGTGGAGCATCGACTACGGCGGCGGACACGTCCATGAACGGCGGACCGGTGGCCCTATTCC	572
580	Qy	AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla	599			



Pred. No.: 0.000243 Length: 748  
 Score: 362.00 Matches: 72  
 Percent Similarity: 83.91% Conservative: 1  
 Best Local Similarity: 82.76% Mismatches: 14  
 Query Match: 7.05% Indels: 0  
 DB: 29 Gaps: 0

US-09-917-376-1 (1-957) x CE342004 (1-748)

Qy 782 GlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSer 801  
 Db 194 GGCCACCTTCACCTTCACCTTCCTTCACCTTCACCTTCACCTTCACCTTCCT 253  
 Qy 802 ProSerProSerProSerProSerProSerProSerProSerProSerProSer 821  
 Db 254 CCTTCACCTTCCTTCACCTTCACCTTCCTTCACCTTCACCTTCCTTCACCTTC 313  
 Qy 822 SerSerProSerProSerProSerProSerProSerProSerProSerProSer 841  
 Db 314 CCTTCACCTTCCTTCACCTTCCTTCACCTTCCTTCACCTTCCTTCACCTTCCT 373  
 Qy 842 AlaSerProSerProSerProSerProSerProSerProSerProSerProSer 861  
 Db 374 CCTTCCTTCACCTTCACCTTCCTTCACCTTCCTTCACCTTCCTTCACCTTCCT 433  
 Qy 862 ProThrProSerSerPro 868  
 Db 434 CCTTCACCTTCACCTTCCTTCCT 454

RESULT 9  
 CF868882 751 bp mRNA linear EST 31-OCT-2003  
 LOCUS trico16xm14.b1 T.reesei mycelial culture, Version 6 October 2003  
 DEFINITION Hypocrea jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION CF868882  
 VERSION  
 KEYWORDS  
 SOURCE

ORGANISM Hypocrea jecorina (anamorph: Trichoderma reesei)

REFERENCE 1 (bases 1 to 751)  
 AUTHORS Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D., Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and Dean,R.A.

TITLE Analysis of the protein processing and secretion pathways in a Trichoderma reesei EST dataset  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Ralph A. Dean  
 Fungal Genomics Laboratory  
 North Carolina State University  
 Campus Box 7251, Raleigh, NC 27695, USA  
 Tel: 919-513-0020  
 Fax: 919-513-0024  
 Email: ralph.dean@ncsu.edu

FEATURES  
 source Location/Qualifiers  
 1..751  
 /organism="Hypocrea jecorina"  
 /mol\_type="mRNA"  
 /strain="QM6a"  
 /db\_xref="taxon:51453"  
 /clone="trico16xm14"  
 /dev\_stage="mycelia"  
 /clone\_lib="T.reesei mycelial culture, Version 6 October 2003"  
 /note="Vector: pREP3Y; Site 1: Not 1/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Alignment Scores:  
 Pred. No.: 0.000255 Length: 751

Score: 361.50 Matches: 88  
 Percent Similarity: 51.61% Conservative: 40  
 Best Local Similarity: 35.48% Mismatches: 103  
 Query Match: 7.04% Indels: 17  
 DB: 14 Gaps: 6

US-09-917-376-1 (1-957) x CF868882 (1-751)

Qy 453 AsnAspLeuThrValTrpAspSerGlyGlyGlnIleHisIleAlaProMetVallySgLy 472  
 Db 15 CACGATTCACCACTGGGACACGGCCACATGTGTCAATCAATCACTGGCAGCGG 74  
 Qy 473 LeuGluThrAlaValAsnAspLeuLeuSerProSerGlyAlaProLeuLeuSer 492  
 Db 75 ATCGAGGAATTCGTCAGGACCTCGCTCTGCACCGGGGGAAGCAGCTATTGGCC 134  
 Qy 493 AlaLeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValPro 509  
 Db 135 GCAGTCGGAGACGACCAACGGCTTCACCTTTCAGCAGAAACGACCTTCGGGACATCGCCG 194  
 Qy 510 SerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTrpAlaGluLeu 529  
 Db 195 CAGACGGTCTGGGACGCCACATGCGGCACCTCGACGAGGTGACATACCGCGGGAAC 254  
 Qy 530 AsnProSerIleIleValAlaGlySerPheAspProSerSerGlnProAsnAspArg 549  
 Db 255 TCGGTCAAGAGGTCGTCGGTCGCAACACCGCGCGGCAACAG----- 302  
 Qy 550 HisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly 559  
 Db 303 ---GTGGCCATCTCGTCGCGCGGGCGCGAGCTGGAGCATCGACTACCGCGCGGACACG 359  
 Qy 570 ValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 589  
 Db 360 TCCATGAACCGCGGCGACGCTGGCTATTTCGCGCGGCGACACGATCTCTGGTGCACC 419  
 Qy 590 GlyAspProGlyGlnProValValTrpAlaValGlyPheGlyAsnSerTrpAlaAlaSer 609  
 Db 420 GCCTCGTCGGG-----GTGACGGCTCGAGTTCAGGGGAGCTTTGCTCTCGTC 470  
 Qy 610 GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyr 629  
 Db 471 TCGAGCCTCGCGCGGGCGCTCATCGCTCGGACAAAGACCAACACAGCGCTTCTTAC 530  
 Qy 630 AlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVal 649  
 Db 531 GCGGCTCCGGATCGACCTTTTACGTACAGCAAGGACACCGCGAGCAGCTTC----- 581  
 Qy 650 AlaAlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValPro 667  
 Db 582 ACGCGCGG---CCAGCTCGGCGAGCGGAGGATCCGGGATATCGCTGCTACCCG 638  
 Qy 668 GlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGly 687  
 Db 639 ACCACCGCGGCGACGTTGTATGTCTCGACCGGCGGCGATATTCGGTCCACAGACTCG 698  
 Qy 688 GlySerSerTrpSerAlaIleThr 695  
 Db 699 GGCACGACCTTTGGCCAAAGTCTCC 722

RESULT 10

CF898982

LOCUS

DEFINITION

trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea

jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 803)

CF898982 803 bp mRNA linear EST 02-JUL-2003  
 trico16xm14 T.reesei mycelial culture, Version 3 april Hypocrea  
 jecorina cDNA clone trico16xm14, mRNA sequence.

ACCESSION

CF898982

VERSION





```

Pred. No.: 0.000305 Length: 791
Score: 360.00 Matches: 71
Percent Similarity: 84.88% Conservative: 2
Best Local Similarity: 82.56% Mismatches: 13
Query Match: 7.01% Indels: 0
DB: 29 Gaps: 0

US-09-917-376-1 (1-957) x CNS028QT (1-791)

QY 783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro 802
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 347 AGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCT 406

QY 803 SerProSerProSerSerProSerProSerProSerProSerProSerProSerSer 822
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 407 AGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCT 466

QY 823 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerAla 842
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 467 AGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCT 526

QY 843 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 862
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 527 AGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCT 586

QY 863 ThrProSerSerSerPro 868
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 587 AGCCCTAGCCCTAGCCCT 604

RESULT 12
CE190185/c
LOCUS
DEFINITION tigr-gss-dog-1700037149947 Dog Library Canis familiaris genomic,
Genomic survey sequence.
ACCESSION CE190185
VERSION CE190185.1 GI:35345838
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
REFERENCE 1 (bases 1 to 522)
AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
Venter,J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
MEDLINE 22875432
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
FEATURES
    source
        Location/Qualifiers
            1..522
                /organism="Canis familiaris"
                /mol_type="genomic DNA"
                /strain="Standard Poodle"
                /db_xref="taxon:9615"
                /clone_lib="Dog Library"
                /note="Site 1: BstXI; Libraries were prepared from
                peripheral blood"
ORIGIN
Alignment Scores:
Pred. No.: 0.0003 Length: 522
Score: 355.00 Matches: 72
Percent Similarity: 86.05% Conservative: 2
Best Local Similarity: 83.72% Mismatches: 8

Pred. No.: 0.000333 Length: 493
Score: 353.00 Matches: 70
Percent Similarity: 83.72% Conservative: 2
Best Local Similarity: 81.40% Mismatches: 14
Query Match: 6.87% Indels: 0
DB: 28 Gaps: 0

Query Match: 6.91% Indels: 4
DB: 29 Gaps: 1

US-09-917-376-1 (1-957) x CE190185 (1-522)

QY 787 SerProSerProSerValSerProSerAlaSerProSerLeuSerProSerProSerPro 806
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 259 TCGCCATCACCATCGCTGCTCCCATCTCCATCGGCATCTTCTCGCCATCACCATCGCCA 200

QY 807 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerSer 822
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 199 TCGCCATCTCATCTCCATCACCATCGGCATCTTCTCGCCATCACCATCACCATCACC 140

QY 823 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerAla 842
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 139 TCGCCATCTCCATCACCATCGGCATCTCCATCACCATCGGCATCGGCATCTCCATCGCCA 80

QY 843 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 862
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 79 TCTCCATCGGCATCTTCTCGCCATCGGCATCGGCATCGGCATCTTCCATCGGCATCTCCA 20

QY 863 ThrProSerSerSerPro 868
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 19 TCGCCATCTTCTCGGCCA 2

RESULT 13
CC066065 493 bp DNA linear GSS 16-APR-2003
LOCUS
DEFINITION CSU-K33r.42E22.77 CSU-K33r Aedes aegypti genomic clone
CSU-K33r.42E22, genomic survey sequence.
ACCESSION CC066065
VERSION CC066065.1 GI:29904571
KEYWORDS GSS.
SOURCE Aedes aegypti (yellow fever mosquito)
ORGANISM Aedes aegypti
REFERENCE 1 (bases 1 to 493)
AUTHORS Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
TITLE End sequencing of Aedes aegypti BACS
JOURNAL Unpublished (2003)
COMMENT Other GSSs: CSU-K33r.42E22.SP6
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
Library was provided by Susan Brown and Dennis Knudson at Colorado
State University.
Seq primer: T7
Class: BAC ends.
FEATURES
    source
        Location/Qualifiers
            1..493
                /organism="Aedes aegypti"
                /mol_type="genomic DNA"
                /strain="Rexville"
                /db_xref="taxon:7159"
                /clone="CSU-K33r.42E22"
                /clone_lib="CSU-K33r"
                /note="Vector: pBeloBAC11; Site_1: HindIII"
ORIGIN
Alignment Scores:
Pred. No.: 0.000333 Length: 493
Score: 353.00 Matches: 70
Percent Similarity: 83.72% Conservative: 2
Best Local Similarity: 81.40% Mismatches: 14
Query Match: 6.87% Indels: 0
DB: 28 Gaps: 0

```



## ORIGIN

**Alignment Scores:**

US-09-917-376-1 (1-957) x AZ374273 (1-259)

Search completed: May 14, 2004, 16:17:46  
Job time : 5637 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 14, 2004, 08:53:04 ; Search time 158 Seconds  
(without alignments)  
3361.316 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MORSNIRLTMRRLVSL.....RASFGSNPATPTVYLOX 957

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cg2\_1/USPIO\_spool/US0917376/runat\_14052004\_085257\_4644/app\_query.fasta\_1.1095  
-DB=issued Patents NA -QFWT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0917376 @CGN 1 1 75 orunat\_14052004\_085257\_4644 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBIOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:\*

1: /cg2\_6/ptodata/2/ina/5A COMB.seq:\*  
2: /cg2\_6/ptodata/2/ina/5B COMB.seq:\*  
3: /cg2\_6/ptodata/2/ina/6A COMB.seq:\*  
4: /cg2\_6/ptodata/2/ina/6B COMB.seq:\*  
5: /cg2\_6/ptodata/2/ina/PCTUS COMB.seq:\*  
6: /cg2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	409	8.0	11707	3	US-09-136-574A-1
2	345	6.7	18596	3	US-09-318-448-11
3	345	6.7	18597	4	US-09-362-665-8
4	345	6.7	18597	4	US-09-363-333-8
5	334	6.5	6416	3	US-09-136-574A-2
6	319.5	6.2	8211	4	US-09-252-991A-13656
7	296.5	5.8	4767	4	US-09-410-551B-28
8	293	5.7	4818	4	US-09-410-551B-32
9	288.5	5.6	4131	4	US-09-252-991A-13773
10	288.5	5.6	4571	4	US-09-410-551B-18
11	288	5.6	288	4	US-09-119-507B-111
12	288	5.6	288	4	US-09-547-693-111

c 13	282.5	5.5	441529	3	US-09-103-840A-1	Sequence 1, Appli
c 14	281.5	5.5	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 15	280	5.5	3129	4	US-09-252-991A-13873	Sequence 13873, A
c 16	279	5.4	4547	4	US-09-410-551B-22	Sequence 22, Appl
c 17	273	5.3	4674	4	US-09-410-551B-26	Sequence 26, Appl
c 18	269	5.2	4478	4	US-09-410-551B-16	Sequence 16, Appl
c 19	266	5.2	4188	4	US-09-252-991A-13774	Sequence 13774, A
c 20	266	5.2	77536	4	US-09-410-551B-1	Sequence 1, Appli
c 21	265.5	5.2	77536	4	US-09-410-551B-1	Sequence 1, Appli
c 22	265	5.2	4737	4	US-09-410-551B-30	Sequence 30, Appl
c 23	260	5.1	3337	1	US-08-072-610-1	Sequence 1, Appli
c 24	260	5.1	3337	2	US-08-719-822B-1	Sequence 1, Appli
c 25	260	5.1	3337	3	US-09-092-458-1	Sequence 1, Appli
c 26	258	5.0	4725	4	US-09-410-551B-24	Sequence 24, Appl
c 27	257	5.0	4466	4	US-09-410-551B-20	Sequence 20, Appl
c 28	257	5.0	12001	1	US-08-458-568A-11	Sequence 11, Appl
c 29	255.5	5.0	3979	4	US-09-865-621A-4	Sequence 4, Appli
c 30	255.5	5.0	4810	4	US-09-865-621A-3	Sequence 3, Appli
c 31	255.5	5.0	5163	4	US-09-865-621A-7	Sequence 7, Appli
c 32	255.5	5.0	5509	4	US-09-865-621A-1	Sequence 1, Appli
c 33	251	4.9	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 34	250	4.9	441529	3	US-09-103-840A-1	Sequence 1, Appli
c 35	248	4.8	1926	4	US-09-249-585A-4	Sequence 4, Appli
c 36	248	4.8	1931	2	US-09-130-114-2	Sequence 2, Appli
c 37	247	4.8	3979	4	US-08-646-242-6	Sequence 6, Appli
c 38	238	4.6	47981	4	US-09-679-279-1	Sequence 1, Appli
c 39	237.5	4.6	25165	4	US-09-453-702B-39	Sequence 39, Appl
c 40	237	4.6	3427	4	US-09-865-621A-8	Sequence 8, Appli
c 41	237	4.6	3518	4	US-09-865-621A-5	Sequence 5, Appli
c 42	237	4.6	4647	4	US-09-252-991A-5730	Sequence 5730, Ap
c 43	234	4.6	6414	4	US-09-134-001C-1626	Sequence 1626, Ap
c 44	232	4.5	13842	3	US-09-105-537-30	Sequence 30, Appl
c 45	232	4.5	36778	3	US-09-105-537-5	Sequence 5, Appli

#### ALIGNMENTS

##### RESULT 1

; Sequence 1, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for

; Treating Cellulose Containing

; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/932,571

; FILING DATE: September 19, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Bak, Mary E.

Fabrics Using Truncated

REGISTRATION NUMBER: 31,215  
 REFERENCE/DOCKET NUMBER: 1997US001/CIP  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 215-540-9200  
 TELEFAX: 215-540-5818  
 TELEX: <Unknown>  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 11707 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-09-136-574A-1

Alignment Scores:  
 Pred. No.: 3,03e-11 Length: 11707  
 Score: 409.00 Matches: 263  
 Percent Similarity: 32.37% Conservative: 163  
 Best Local Similarity: 19.98% Mismatches: 386  
 Query Match: 7.96% Indels: 504  
 DB: 3 Gaps: 58

US-09-917-376-1 (1-957) x US-09-136-574A-1 (1-11707)

QY	14	ArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGly	33
DB	735	AAGAGGGTAATTCATTCCTTTCCTTATTGTTTTTTTATAAACACGCTTGAGGT	791
QY	34	ValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThrThrGlnProTyrThr	53
DB	792	ACTTTG-----ATATTCATCAGGAAGCAAAAGCAGCAGCA-----TATACT	833
QY	54	TrpSerAsnValAlaIleGlyGlyGlyGlyPheValAspGlyIleValPheAsnGluGly	73
DB	834	-----GTGATTTTGAAGGT	848
QY	74	AlaProGlyIleLeuTyr-----ValArgThrAspIleGlyGly	86
DB	849	GCTGATCTTATCTTCTTCTGTTATGGAATCGACATAGCATGTTGACATGGGCAAT	908
QY	87	MetTyrArgTrpAspAlaAlaAsnGlyArgTyrIleProLeuLeuAspTrpValGlyTrp	106
DB	908	-----	908
QY	107	AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys	126
DB	909	-----GCATATAATGTAAAGTAGTGTCTAGGGGTGTC-----AATAGAAGTTCA	953
QY	127	ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu	146
DB	954	ATATGGATGGATGTCAGTT-----GACGTATAA	983
QY	147	ArgSerSerAspGlnGlyAlaThrTrpGlnIleThr-----	158
DB	984	AACATTATGAACAATGGAACCAACATCGGTAGTTTCAGCGTATGATAAACATAGCTACCAG	1043
QY	159	ProLeuProPheLysLeu-----GlyGlyAsnMetProGlyArg	171
DB	1044	AGCCGGTTCATTTGGTATCTCAGCGGTTTACGACGATGGAAGTGGGGTTAAGAGTACT	1103
QY	172	GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla	191
DB	1104	CTCATAGGTAGGTGGTGGCTATT-----CCAAAT	1133
QY	192	ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr	211
DB	1134	-----TATTCGAAGAAAATT---	1148
QY	212	AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer	231
DB	1149	-----GTTGGTAAATGCAATGCTCCAAATATTATTAAGCAATGTCAGGAATTTGTTAATT	1196

QY	232	AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerSerLeuGlnAla	251
DB	1197	GTAATACACACAATTTGTAAAGCGAAGTAGATATATATTTGTAC-----	1241
QY	252	SerLysThrIlePheValGlyValAlaAspProAsn-----AsnProValPhe	267
DB	1242	-----TATATCCAAATATATGATGATATATAGTTACCTATCAATATCAGTCAGCA	1289
QY	268	TrpSerArg-----AspGlyGlyValAlaThr-----TrpGlnAla-----Val	279
DB	1290	TTTTCAAGTGGATTTGAAAGTGGCACTACCGAGGTTTGCAGGCAAGGGAAGCGGTGT	1349
QY	280	ProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisVal	299
DB	1350	ACGTAACACCATAGCGTTGTGTCATATAGTGGCAAGTATAGT-----	1394
QY	300	LeuTyrIleAla-----ThrSerAsnThrGlyGly-----ProTyrAspGly---	313
DB	1395	TTGTACGTCAAGTGAAGAACGTCAAATTTGGCATGTCACAGATTCGGGTAGATACAATT	1454
QY	314	SerSerGlyAspValTrpLysPheSerVal-----ThrSerGlyThrTrp	328
DB	1455	TTGGAACAGGTAAGTGTATAAATAAGTGTTTGGGTTTATCAGAACAGTGTTCACACT	1514
QY	329	ThrArgIleSerProValProSerThrAspThrAlaAsnAsp-----TyrPheGlyTyrSer	347
DB	1515	CAAAAAATGATTAATATGCAAAAGAAAGATTGCTACAGATCCTTCAACAAGCTATGAA	1574
QY	348	GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp	367
DB	1575	AATCTGATATATAACAGCGGATGTACCAGATTAATACGTGGTTGAGCTAGTGAAGCTAC	1634
QY	368	TrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp	387
DB	1635	-----TCAATTCCTGCTGGTGTACAGTTAGCGAGTTGTTG	1670
QY	388	AspTrpTrpSerTyrProAsnArgSerLeuArgTyr-----ValLeu	401
DB	1671	CTTTATGTTGAGGCACAAAATGCAAAATTTGGCTTCTCGGTTGATGATTTAAAGATTAT	1730
QY	402	AspIleSer-----AlaGluPro-----	407
DB	1731	GATTTATCAAGTTGGCTGAACCTGAATGGGAGATACCATCTTTGATAGAAAAGTATAGA	1790
QY	407	-----	407
DB	1791	GATTTATTTCAAAGTAGGATAGCTTTGTCTTACAAAACATTCCTCTGATACAGAAAAG	1850
QY	407	-----	407
DB	1851	AAGATGGTTTTGAAGCATTTTCAATAGTATTACTGCAGGGAACGAAATGAAACCATCAGAG	1910
QY	408	-----TrpLeuThrPheGly	412
DB	1911	TTACTTGTGATGAAAAATACCTTACAACCTTTAGCAAGCAGACGAATTTGTAAATTTGCA	1970
QY	413	ValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAla-----	430
DB	1971	ACAAGTAACACATGTCATGCCATCAGGTCATACACTGGTTTGGCATGAGCAACACCCGAC	2030
QY	430	-----	430
DB	2031	TGGTTTTTCAAGGACACAAATGGAATACGTTTGACAGGATGCATTGCTAAGCAGATTTA	2090
QY	430	-----	430
DB	2091	AAACAGTATATTATACGTTAGTGGGAAGATATAAAGGGAAGGTTTATGCTAGGATGTG	2150
QY	431	-----MetAlaIleAspProPheAsnSerAspArgMetLeu-----	442
DB	2151	GTAATGAAGCAATAGATGAAGTCAAGGTGATGATTCAGGAGATCTAACTGGTACAAAC	2210
QY	443	-----TyrGlyThrGlyAlaThrLeuTyrAla-----	451





```
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stenroos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318,448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 18596
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-318-448-11

Alignment Scores:
Pred. No.: 7,74e-08 Length: 18596
Score: 345.00 Matches: 71
Percent Similarity: 83.91% Conservative: 2
Best Local Similarity: 81.61% Mismatches: 13
Query Match: 6.72% Indels: 1
DB: 3 Gaps: 0

US-09-917-376-1 (1-957) x US-09-318-448-11 (1-18596)
Qy 783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro 802
Db 10388 TCACCATCACCATTCTCCATCCATCACCATCACCATCACCATCACCATCACCACCA 10329
Qy 803 SerProSerProSerProSerProSerProSerProSerProSerProSerProSer 822
Db 10328 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10269
Qy 823 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerAla 842
Db 10268 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10209
Qy 843 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPr 862
Db 10208 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCC 10149
Qy 862 oThrProSerSerPro 868
Db 10148 ATCTCCATCACCATCACCACCA 10130

RESULT 3
US-09-962-665-8/c
; Sequence 8, Application US/09962665
; Patent No. 6537759
; GENERAL INFORMATION:
; APPLICANT: Stanton, Jr., Vincent P.
; TITLE OF INVENTION: POLYPOLYGLUTAMATE SYNTHETASE GENE SEQUENCE
; TITLE OF INVENTION: VARIANCES HAVING UTILITY IN DETERMINING THE
; TITLE OF INVENTION: TREATMENT OF DISEASE
; FILE REFERENCE: 11926-015004
; CURRENT APPLICATION NUMBER: US/09/962,665
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 09/658,659
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 09/596,033
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 09/357,743
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 09/357,024
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: 60/093,484
; PRIOR FILING DATE: 1998-07-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 18597

; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: 701_13751
; OTHER INFORMATION: n = c or a
; NAME/KEY: misc feature
; LOCATION: 716_1293, 2401_2429, 2618_3083, 3125_3635, 4256_4898,
; LOCATION: 5062_5167, 11069_13298, 14479_14730, 14796_15344, 15450,
; LOCATION: 15503_15590, 15840_16149
; OTHER INFORMATION: n = a or g
; NAME/KEY: misc feature
; LOCATION: 732_1379, 1590_2488, 3212_5006, 11238_11422, 11686,
; LOCATION: 12598_13171, 13645_13782, 13806_13813, 14586_14788,
; LOCATION: 15042_15546, 15770
; OTHER INFORMATION: n = c or t
; NAME/KEY: misc feature
; LOCATION: 1322_1688
; OTHER INFORMATION: n = c or g
; NAME/KEY: misc feature
; LOCATION: 2594_11293, 16199_16203
; OTHER INFORMATION: n = g or t
; NAME/KEY: misc feature
; LOCATION: 3619
; OTHER INFORMATION: n = a or t
; NAME/KEY: misc feature
; LOCATION: 14547
; OTHER INFORMATION: nucleotide in position 14547 is t, or absent
US-09-962-665-8

Alignment Scores:
Pred. No.: 7,74e-08 Length: 18597
Score: 345.00 Matches: 71
Percent Similarity: 83.91% Conservative: 2
Best Local Similarity: 81.61% Mismatches: 13
Query Match: 6.72% Indels: 1
DB: 4 Gaps: 0

US-09-917-376-1 (1-957) x US-09-962-665-8 (1-18597)
Qy 783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro 802
Db 10388 TCACCATCACCATTCTCCATCCATCACCATCACCATCACCATCACCATCACCACCA 10329
Qy 803 SerProSerProSerProSerProSerProSerProSerProSerProSerProSer 822
Db 10328 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10269
Qy 823 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerAla 842
Db 10268 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCA 10209
Qy 843 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPr 862
Db 10208 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCACCC 10149
Qy 862 oThrProSerSerPro 868
Db 10148 ATCTCCATCACCATCACCACCA 10130

RESULT 4
US-09-963-333-8/c
; Sequence 8, Application US/09963333
; Patent No. 6664062
; GENERAL INFORMATION:
; APPLICANT: Stanton, Jr., Vincent P.
; TITLE OF INVENTION: THYMIDINE SYNTHASE GENE SEQUENCE VARIANCES
; TITLE OF INVENTION: HAVING UTILITY IN DETERMINING THE TREATMENT
; TITLE OF INVENTION: OF DISEASE
; FILE REFERENCE: 11926-015002
; CURRENT APPLICATION NUMBER: US/09/963,333
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 09/658,659
```

PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 09/596,033  
PRIOR FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: 09/357,743  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 09/357,024  
PRIOR FILING DATE: 1999-07-19  
PRIOR APPLICATION NUMBER: 60/093,484  
PRIOR FILING DATE: 1998-07-20  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 8  
LENGTH: 18597  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 701, 13751  
NAME/KEY: misc feature  
LOCATION: 716, 1293, 2401, 2429, 2618, 3083, 3125, 3635, 4256, 4898,  
LOCATION: 5062, 5167, 11069, 13298, 14479, 14730, 14796, 15344, 15450,  
LOCATION: 15503, 15590, 15840, 16149  
OTHER INFORMATION: n = a or g  
NAME/KEY: misc feature  
LOCATION: 732, 1379, 1590, 2488, 3212, 5006, 11238, 11422, 11686,  
LOCATION: 12598, 13171, 13645, 13782, 13806, 13813, 14586, 14788,  
LOCATION: 15042, 15546, 15770  
OTHER INFORMATION: n = c or t  
NAME/KEY: misc feature  
LOCATION: 1322, 1688  
OTHER INFORMATION: n = c or g  
NAME/KEY: misc feature  
LOCATION: 2594, 11293, 16199, 16203  
OTHER INFORMATION: n = g or t  
NAME/KEY: misc feature  
LOCATION: 3619  
OTHER INFORMATION: n = a or t  
NAME/KEY: misc feature  
LOCATION: 14547  
OTHER INFORMATION: nucleotide in position 14547 is t, or absent  
US-09-963-333-8

Alignment Scores:  
Pred. No.: 7.74e-08 Length: 18597  
Score: 345.00 Matches: 71  
Percent Similarity: 83.91% Conservative: 2  
Best Local Similarity: 81.61% Mismatches: 13  
Query Match: 6.72% Indels: 1  
DB: 4 Gaps: 0

US-09-917-376-1 (1-957) x US-09-963-333-8 (1-18597)

Qy 783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro 802  
Db 10388 TCACCATCACCATTCTCCATCACCATCACCATCACCATCACCATCACCATCACC 10329  
Qy 803 SerProSerProSerProSerProSerProSerProSerProSerProSerProSer 822  
Db 10328 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATC 10269  
Qy 823 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerAla 842  
Db 10268 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATC 10209  
Qy 843 SerProSerProSer-SerProSerProSerProSerProSerProSerProSerProSer 862  
Db 10208 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATC 10149  
Qy 862 oThrProSerSerProSerProSerProSerProSerProSerProSerProSerProSer 868  
Db 10148 ATTCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATC 10130

RESULT 5  
US-09-136-574A-2  
Sequence 2, Application US/09136574A  
Patent No. 6294366  
GENERAL INFORMATION:  
APPLICANT: Farrington, Graham K.  
Anderson, Paige  
Gibbs, Moreland  
Bergquist, Peter  
Daniels, Roy  
Morgan, Hugh W.  
Williams, Diane P.  
TITLE OF INVENTION: Compositions and Methods for  
Treating Cellulose Containing  
Cellulase Enzyme Compositions  
NUMBER OF SEQUENCES: 49  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Center, P.O. Box 457  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/136,574A  
FILING DATE: 19-Aug-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/932,571  
FILING DATE: September 19, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: 1997US001/CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6416 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-136-574A-2

Alignment Scores:  
Pred. No.: 8.13e-08 Length: 6416  
Score: 334.00 Matches: 219  
Percent Similarity: 33.27% Conservative: 125  
Best Local Similarity: 21.18% Mismatches: 314  
Query Match: 6.50% Indels: 376  
DB: 3 Gaps: 56

US-09-917-376-1 (1-957) x US-09-136-574A-2 (1-6416)

Qy 18 SerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProfile 37  
Db 1189 TCTGTATGCGCAGAAACA-----GCTGCATCATTGCGGTGCTTCAGTT 1233  
Qy 38 AlaIleThrAlaSerProAlaHisAlaAla----- 47  
Db 1234 GTATAAAGGAAGAAATCTCAGAAAGCAGCTTTATCTCAACATGCCAAGACCTG 1293  
Qy 48 -----ThrThrGlnPro-----TyrThrTrpSerAsnValAlaIle 59  
Db 1294 TTTGAATTGCGGATACCAAGAGAGTATGCGGGTATCTGCTGCAACAGGTTCTTCTAC 1353



```

Db 3034 TTCAGAGCATGACGAAT-----TATGGAGAGAAATCGCAGGTGACGCTG 3078
Qy 714 TyrProAlaValPheValValGlyThrIleGlyValThrGlyAlaTyrArgSerAsp 733
Db 3079 TAT-----GTAGATGGT----- 3090
Qy 734 AspCysGlyThrThrTrpValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrpGly 753
Db 3091 -----GTTCTGGTA-----TGGGG 3105
Qy 754 GlnAlaIleThrClyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArg 773
Db 3106 CAG----- 3108
Qy 774 GlyIleValTyrGlyAspIleGlyAlaProSerGlySerProSerProSerValSer 793
Db 3109 -----GAGCGGGAGGAGCG----- 3123
Qy 794 ProSerAlaSerProSerLeuSerProSerProSerProSerProSerProSer 813
Db 3124 -----ACACCTGTCACCGACAAGCAGCAGCAACACCAACG 3156
Qy 814 ProSerProSerSerProSerProSerProSerProSerProSerProSerProSer 833
Db 3157 CCAACTCCGACAGCAACCCCAACACCTACACCTACACCGACCCCGACACGACAGTGA 3216
Qy 834 ProSerArgSerProSerProSerAlaSerProSerProSerProSerProSerProSer 853
Db 3217 GCACCCCAACACCGCAGCGGATCACCAGTGGTGGCAGTACTGGAGCGCGAGT 3276
Qy 854 SerSerProSerSerProSerProSerProSerProSerProSerProSerValSer 873
Db 3277 GAGAGT-----TACGGTGGCGCTG 3294
Qy 874 LysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeu 893
Db 3295 AAGGTATGTATCGCAATGGAATTTAAGCAGCCCGACGAATGATTGAATCCTAAGATA 3354
Qy 894 GlnValValAsnThrGlySerSerSerValAspLeuSerThrValThrValArgTyrTrp 913
Db 3355 AAGATAGAGAATTTGGGACGACAGCGGTAGATCTTAGCAGGGTGAAGGTAAAGTACTGG 3414
Qy 914 PheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrpAlaAlaIleGly 933
Db 3415 TACACATAGATGGTGGGACGACAGAGTGTA----- 3447
Qy 934 CysGlyAsnIleArgAlaSerPheGlySerValAsnProAla 947
Db 3448 -----AGTGTAGCAGCAGCATAAATCCTGCG 3474

```

## RESULT 6

```

US-09-252-991A-13656
; Sequence 13656, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13656
; LENGTH: 8211
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13656

```

Alignment Scores:

```

Pred. No.: 5,65e-07 Length: 8211
Score: 319.50 Matches: 290
Percent Similarity: 32.56% Conservative: 133
Best Local Similarity: 22.32% Mismatches: 449
Query Match: 6.22% Indels: 432
DB: 4 Gaps: 69

US-09-917-376-1 (1-957) x US-09-252-991A-13656 (1-8211)

Qy 18 SerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIle 37
Db 3326 GCAATGGCAGCAGCTCAGCGGCACTGCGGAAACCGGGCAGCAGCGTGATCTCACCAGC 3385
Qy 38 AlaIleThr-----AlaSerProAlaHisAlaAlaThrThrGlnPro----- 51
Db 3386 GCAACGGCAATCCGATCGCCGAGGTCAACCGCGAGCGCAGCGCACTGCACCTACACCC 3445
Qy 52 -----TyrThrTrpSerAsnVal-----AlaIleGly 60
Db 3446 CGTCCAGCGCGATCGCCAAACGGTACTGTGTC-AACGTGTGTGGCGGAGACGCCGCGT 3504
Qy 61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80
Db 3505 AACAGCAGCGCGCGGAGCGGTGACCGTGCATTCCAGCGCGCGCGCGCGCGGTGATC 3564
Qy 81 ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu 100
Db 3565 AACCGGAGCAACGGCGTGTCTATCAGCGGCACCGCGGAGCGCGTGCACCGTGACCTC 3624
Qy 101 LeuAspTrpValGlyTrpAsn-----AsnTrp 109
Db 3625 ACCGAT---GCCGCGGCAACCCGATAGCGGAGGTCAACCGCGAGCGGAGCGCACTCG 3681
Qy 110 GlyTyr-----AsnGly-----ValValSerIleAlaAlaAsp 120
Db 3682 AGCTTTCAGCGCGGACGCGCGCCCAACCGCGCGGTGATCGTCCACGCGCACCGAC 3741
Qy 121 Profile---AsnThrAsnLysValTrpAla-----AlaValGlyMetTyrThr 135
Db 3742 CCGACCGGCAATACCGCGCGCGCGCGCGCGCGCGCGCGCGGTGAGCGCGTGGCGCGCG 3801
Qy 136 AsnSerTrpAspProAsnAspGlyAlaIleLeu----- 146
Db 3802 CCGGTGATCGATCGAGCAACGGCACGACCATCAGCGCGACCGCGGAGCGCGGCGCAAG 3861
Qy 147 -----ArgSerSerAspGlnGlyAla 153
Db 3862 GTGATCTCACCAGCGCAACCGCAACCCGATCGCGGCAACCCACCGCGAGCGCGCGC 3921
Qy 154 ThrTrpGlnIleThrPro---LeuProPheLysLeuGly----- 165
Db 3922 AACTGGACCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGGTGAGCGCGTGGCGCG 3981
Qy 166 -----GlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspPro 181
Db 3982 CAGACCCCTCGGGCAATACCGCGCGCGCGCGCGCGCGCGCGCGCGGTGAGCGCGTGG 4041
Qy 182 Asn-----AsnAspAsnIleLeuTyrPheGlyAlaProSerGly 194
Db 4042 AACACGCTGTGGTCAATCCGAGCAACGGCAACCTGCTCAACGGTACCGCGGAGCGCG 4101
Qy 195 LysGlyLeu-----TrpArgSerThrAspSer 203
Db 4102 AGCACCGGTGACCTTTCAGCGCGCGCAACCGCATCGCGCGAGCACACCGCGCGGTG 4161
Qy 204 GlyAlaThrTrp-----SerGlnMetThrAsnPheProAspValGlyThr 218
Db 4162 AGCGGCAACTGGAGCTTCACCGCGCGGTCTCGCACTACCCAAAC-----GGCACC 4209
Qy 219 TyrIle---AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVal 237
Db 4210 GTGGTCAACGTGACCGCGGAGCGACGCGCGCGCGCAATACCAAC-----GCTCCCGCT 4260

```

QY 238 TrpValAlaPheAspLysSerSerSerLeuGlyGln-----AlaSerLysThr 254  
DB 4261 ACCACGACGGTGTCTCTCGTGGCGTGGATCCCGCAGCAACGAGTTCG 4320  
QY 255 IlePheValGlyValAlaAspProAsnAsnProValPhe----- 267  
DB 4321 GTGATCAGCGGCACCGCAGCGCGGCAACACCATCATCATCATCCGATGGCAACGGCAAC 4380  
QY 268 -----TrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro 283  
DB 4381 CCGATTGGCCAGGTACCCCGCAGCGGTAATCTGGTCTTCACTCCAGGCATCCCG 4440  
QY 283 ----- 283  
DB 4441 CTGCGGATGGCACGGTGTCAACGTGTGGCGCGCAGCCCAAGCAATGTCCAGATGGC 4500  
QY 284 -----ThrGlyPheIleProHisIleGlyValPheAspProValAsn 297  
DB 4501 CCGGGCGGTGATCAGTGTGATGGCGGCGCGCGCGCGCGGTGATCGATCCGAGCAAC 4560  
QY 298 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr-----AspGlySer 314  
DB 4561 GGCACCGAGTAAGCGGTACCGGAGGCGCGCGCGCGGTGATCTCACCAGTGGCGG 4620  
QY 315 SerGlyAspValTrpLysPheSerValThr---SerGlyThrTrpThrArgIleSerPro 333  
DB 4621 GGCACACCGATCGCGCAGCGCACCGCGCGCAGCGGCGCAACTCGAGCTTCACCCCGGCG 4680  
QY 334 ValPro-----SerThrAspThrAlaAsnAspTyr 343  
DB 4681 ACCCGCTGGCCAAACCGCACCGGTGATCAACCGCGTGGCCCGCAGCACCGCGCGCAATACC 4740  
QY 344 PheGlyTyrSerGlyLeuThrIleAspArgGlnHisPro----- 356  
DB 4741 AGCGGTCCGCGCAGCGTACCGTGCATCGCGCGCGCGCGCGGTGATCAATCCG 4800  
QY 357 ---AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArg 375  
DB 4801 AGCAACCGGAGTGTGTCATCAGCGGTACGGCGGAAGCGCGGCCACCGGTGATCTCTC----- 4854  
QY 376 SerThrAspGlyGly-----AlaThrTrpThrArgIleTrpAspTrp 389  
DB 4855 ---ACCGACGGCAACCGCACCGATCGCGCACCGGTACACCGCGCAGCGCAAGTGG 4911  
QY 390 Thr-SerTyrProAsnArg-----SerLeuArgTyrValLeuAs 402  
DB 4912 GCTTTCACCGCCGCCACGCGGTGGCAATGGCAGCGGTGATCAATGCGCTGG----- 4963  
QY 402 pIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValProSerPr 422  
DB 4964 ---CCAGGACGCGCGCGCAACACAGCAGTCT-----CCACGACGCCA---CC 5007  
QY 422 oLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLe 442  
DB 5008 GTGACTCGTGGCGCCAGCA-GCCCGCGTGTGATCGATCCGAGCAACGGTATG---GTGAT 5063  
QY 442 uTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAs 459  
DB 5064 CGCGGTACCGCGAGGTGTGTCCAGGTGATCTCACC----- 5105  
QY 459 pSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAs 479  
DB 5106 CGGCAACGGCAAC-----CCGATCGCGCAGGTACCGCGCGATGGCAGCGGCA 5153  
QY 479 nAspLeuIleSerProProSerGlyAlaProLeu-----IleSerAlaLe 494  
DB 5154 CTGAGCTTCAGGCC-----GGCAGCGCGTGTCCAATGGCAGCGGTGATCAATGCGGT 5207  
QY 494 uGly---AspLeuGlyGlyPheThrHisAlaAspVal---ThrAlaValProSerThrI 512  
DB 5208 GGCCCGAGGACGTGCGCGCAACACCCAGCGCGCGCGCGCAGCACCGGTGACTGGGTGGC 5267  
QY 512 ePheThrSerProValPhe-----ThrThrGlyThrSerVal---AspTyrAlaGluLe 529

DB 5268 GCCGCGCGCGCGGTGATCGACCGCAACCGCAGCGGTGATCGCGGTACCGCGCAACGC 5327  
QY 529 uAsnProSerIleIleValArgAlaGlySerPheAspProSerSerSerGlnProAsnAspAr 549  
DB 5328 CGGTGCGACGGTGTCTCACCAGTGGCGGCGCAACCGCATCGCGCCAGGCC----- 5379  
QY 549 gHisValAlaPheSerThrAspGlyGlyLysAsnTrp----- 561  
DB 5380 ---ACCGCCGATGGCAGCGGCAACTGGAGCTTCACCCCGGGCAGCGCGCT 5426  
QY 562 -----PheGlnGlySerGluProGlyGlyValThrThrGlyGln 574  
DB 5427 GGCCAAACGGCACCGGTGATCAATCGGTGGCCCGCAGGATCCGCGCGCAATACCGCGGCC 5486  
QY 574 YThr----- 575  
DB 5487 GACCAGCACCGGTGGACGGGTGGCCCGCGCACCCCGGTGGTCAACCCGAGCAACGC 5546  
QY 576 -----ValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGl 593  
DB 5547 CAGCGTGTGTCGCGGTACCGCGGAAGCGCGGCCACCGGTGATCTCACCAGCGGGCGG 5606  
QY 593 yGlnProVal-----ValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGln 611  
DB 5607 CAACCCGATCGCGCAGGTTCACCGCGCAGCGCGGCAAC---TGGAGCTTCACGCGCGG 5663  
QY 611 yValPro---AlaAsnAlaGlnIleArgSer----- 620  
DB 5664 CAGCCGCTGGCCAAACCGGTGATCAATGCGTGGCCCGCAGGACCGCGCGCAACAC 5723  
QY 621 -----AspArgValAsnProLysThr---PheTyrAlaLe 631  
DB 5724 CAGCGCGCGCCAGCACCGGTGACTCGGTAGCCCGCGCCACCCCGGTGCTCGATCC 5783  
QY 631 uSerAsnGlyThrPheTyrArgSer-----ThrAs 641  
DB 5784 GAGCAACCGTACCGGTGATCAGCGGTACCGCGAAGCCCGGGCCACCGGTGATCTCACC 5843  
QY 641 pGlyGlyValThrPheGlnProValAlaAlaGlyLeu-ProSerSerGlyAlaValGlyV 661  
DB 5844 CGGCGGCG---GGCAACCGATACCGCAGCGCCACCCCGATGGCAGCGGCAACTGGAG 5897  
QY 661 alMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyL 681  
DB 5898 CTTCACTCCGGGCACA-CGCTGACCAACGCGCATGATCAATGCGTGGCCAGGACG 5956  
QY 681 euTyrHisSerThrAsnGly-----GlySerSerTrpSerAlaIleThrGlyValSerS 699  
DB 5957 CCGCGCGCAACACCGCGGTCCGCTCAGCACACAGTGGACCGGTGGCGCGGCCACCC 6016  
QY 699 erAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheV 719  
DB 6017 CGGTGATCGCCCGGAGCATGGTGTCAAACTCAGCGGCACCGCGCAACCCCGGTCGGG 6076  
QY 719 alValGlyThrIleGlyGly----- 725  
DB 6077 TGATCTCACCAGTGGCAATGGCAACCGATCGCGCCAGACCCCTCGCGCAGGTAGCGCA 6136  
QY 726 -----ValThrGlyAlaTyrArgS 732  
DB 6137 ACTGGACCTTCACACCGGGCACCGCTGGCGCAACCGCAGCGGTGTGTCAACGCGGTGCC 6196  
QY 732 erAspAspCysGlyThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnT 752  
DB 6197 AGGACCCCGCGCGCAATACACAG-----CGTCCGCGCCAGCACCGGTGGATACG 6246  
QY 752 rpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg-ValTyrIleGlyThrAsn 771  
DB 6247 GTGCGCGCGCCAGCGCGGTGATCAATGCCAGCAACGCGGTGATCACCAGCGCACCGC 6306  
QY 772 -----GlyArgGlyIleValTyrGlyAspIleGlyAlaProSerGlySer-ProSe 789

```
Db 6307 GAGTGGCGCCAAAGTATCTCACCAGCGGCAACCGCAACCGATCGCGGAGACCACC 6366
Qy 789 rProSerVal-----SerProSerAlaSer-----ProSerLeu----- 800
Db 6367 GCCAGCGCAGTGGCAACTGACCTTCAACCCCGGCACCGCGCTGGCCAAACGGTACGGTG 6426
Qy 801 -----SerProSerPro-----SerProSerProSerProSerPro----- 816
Db 6427 ATCAACCGCGTGGCGGAAGACCGCGCGGCAACCGCAGCTCGCGGCAGCACCGCGTG 6486
Qy 817 -----SerSerSerProSerProSerProSerProSerProSerProSerPr 826
Db 6487 GACTCGTGGCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTGACC 6546
Qy 826 oSerPro-----SerProSerProSerProSerProSerProSerProSerAr 836
Db 6547 GGCACCGCGGACCGGACAGTACAGTGGTATCGTGTCAACGGCGACACCGCCAAACCGG 6606
Qy 836 gSerProSerProSerAlaSerProSerProSerProSerProSerPro-----Se 849
Db 6607 ATCAGCGTCAACCGTCAAGCGCGCGCGGCAACTTCAAGCTCGCTCGCGCGCGCGCTGATC 6666
Qy 849 rProSer-----ProSerSerProSerProSerProSerProSerProSerSe 859
Db 6667 ACCGCGCACTGATCGCGCGGTTGCGGTGACGCGCGCGCGCAACGTCAGCGGCGCGCC 6726
Qy 859 rProSer-----ProThrPr 864
Db 6727 ACCATCAACCGCGCGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGATACC 6786
Qy 864 oSerSerSerProValSerGlyValValValValValValValValValValValValPr 884
Db 6787 TGGATCAACCGCGCGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTC- 6825
Qy 884 oGlyAspAsnGlnIleLysProGlyLeuGlnValValValValValValValValValVal 904
Db 6826 ---GATGTACGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG- 6858
Qy 904 pLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuVa 924
Db 6859 ---CAGTGGTCAACGCGTCAAGTTC---GCCGGCGAAGACGGCTACGAGCGCGAGGTGACG 6911
Qy 924 lTyrAsnCysAspTrpAlaAlaIleGlyCysGlyValValValValValValValValVal 941
Db 6912 CCATACCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCC 6971
Qy 942 -GlySerValAsnPro-----AlaThrProThrAlaAsp 952
Db 6972 CGCGCGCATGGCGCGCGTTCGCGGAGGCGCGCTCGACCGTCAACCGCGCGAC 7020
```

## RESULT 7

```
US-09-410-551B-28
; Sequence 28, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYPEPTIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
```

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 4767
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKs synthase fragment
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(4766)
US-09-410-551B-28
```

## Alignment Scores:

```
Pred. No.: 4,23e-06 Length: 4767
Score: 296.50 Matches: 233
Percent Similarity: 35.16% Conservatives: 101
Best Local Similarity: 24.53% Mismatches: 397
Query Match: 5.77% Indels: 223
DB: 4 Gaps: 46
```

US-09-917-376-1 (11-957) x US-09-410-551B-28 (1-4767)

```
Qy 9 LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28
Db 1827 CTGACCGAACAACGAGACCGGCTGCGCGCTACCTGCGCGGCTGCGCGCGCGGATATA 1886
Qy 29 AlaAlaLeuGlyValLeuProIleAlaThrAlaSerProAlaHisAlaAlaThr 48
Db 1887 CGGCTGTGGATCGACGCTG-----CGGCTGACACGCTCGGT-GTTCGAGCAGCCGCGC 1939
Qy 49 ThrGlnProTyr----- 52
Db 1940 GGTACTCTTCGAGATGACACGCTCAGCGGACCGCGGTGACCGACCCGAGATCGTGT 1999
Qy 53 -----ThrTrpSerAsnValAlaIleGlyGlyGly-----PheVal 65
Db 2000 TGTCTTTCCCGCGGAGGCTGCGAGTGGCTGGGATGGGAGTGCACGTGCGGATTCGTC 2059
Qy 66 AspGlyIleValPheAsnGlyGly-----AlaProGlyIleLeuTyrValArgThrAsp 83
Db 2060 GGTGTGTTCGCGGAGCGGATGGCGGAGTGTGCGCGGCGGTTCGCGAGATTCGT---GGA 2116
Qy 84 IleGlyGlyMetTyrArgTrpAspAlaAlaAlaGlyArgTrpIleProLeuLeuAspTrp 103
Db 2117 CTGGATCTGTTCAGGTTCTGGATGATCCGCGCGGTGG-TGGACCGGGTTGAT---TGG 2172
Qy 104 ValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsn 123
Db 2173 TCCAGCCC-GCTTCTGGCGATG-----ATGGTTTCTTCGCGCGCG- 2213
Qy 124 ThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSer---TrpAspProAsnAsp 142
Db 2214 -----GTGTGCGAGCGCGCGCTGTCGCGCGCGATCGGTCATCGGCATTCGCGAG 2264
Qy 143 GlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGln----- 156
Db 2265 GGTGAGATCGCCGAGCTTGTGTGGCGGGTGGGTGTCATACGCGATGCCCGCGGATC 2324
Qy 157 IleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArg 176
Db 2325 GTGACCTTGGCAGCCAGCGATCGCCCGGGCGCTGGCGGGCGCGGCGGATGCATCC 2384
Qy 177 LeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 196
Db 2385 GTCGCCCTG---CCCGCGCAGGATGTCGAGTGGTGGCGAGCGGGCC- 2426
Qy 197 LeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVal 216
Db 2427 ---TGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2483
Qy 217 GlyThrTyrIleAlaAsnProThrAspThrThrGly----- 228
```

2484 GTGACCATGCTCTCACCCTCATGAGGACAAAGGGTGGGGTGGCGGATCACCGTC 2543  
Qy 229 ---TyrGlnSerAspIleGlnGlyValValTrpVal-----AlaPheAspLysSer 244  
Db 2544 GACTATGCTCTGCACACCCCGCAGCTCGAGCTGATCGCGACGAACACTCTCATCACT 2603  
Qy 245 SerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsn 264  
Db 2604 AGGCACAGCAGCTCGCAGACCCCGCTCGTCCGCTGGCTGTCGACCCGTCGACGCCACCC 2660  
Qy 265 ProValPheTrp-----SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAla 282  
Db 2661 -----TGGTGTGACACCCCGCTGACAGCTGTCAGGCCCGCAGGCGACACCGTGTGTC 2711  
Qy 283 ProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTrpIle 302  
Db 2712 CCGTGGTGTTCACACCCCGCTGACAGCTGTCAGGCCCGCAGGCGACACCGTGTGTC 2771  
Qy 303 AlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSer 322  
Db 2772 GAGTCAAGCCGACCGCGGTGTGTCAGCGCATGACGACGATGTCACCGTGTGCC 2831  
Qy 323 Val-----ThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThr 339  
Db 2832 ACGTGTGTGTGACGACGCGCGACCGCCCGCGATG-----CTCACCGCCCTG 2879  
Qy 340 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 359  
Db 2880 GCACAGCCCTAT-----GTCCACGGGTCTCACCGTCGAC----- 2912  
Qy 360 MetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly 379  
Db 2913 -----TGCCCGCCATCCTC----- 2927  
Qy 380 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 399  
Db 2928 GGCACCAACACACCCGGTACTGGACCTTCCGACCTTCCCAACACACCGGTAC 2987  
Qy 400 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal 419  
Db 2988 -----TGCTC-----GAGTCGGTCTCCCGCGCC 3011  
Qy 420 ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 439  
Db 3012 ACGGCC-----GACTCGGCC 3026  
Qy 440 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 459  
Db 3027 CACCCCGTCTCGGCACCGAGTCCGCTCGCGGG----- 3062  
Qy 460 SerGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAla---Val 478  
Db 3063 TGGCGGCGCGGTGTTACGGTTCCTGCGCGCGGTGGGACCGCGGTGTTATC 3122  
Qy 479 AsnAsnLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly 498  
Db 3123 GCCGAACCTGGCTCGCGCGCGCGCGCCAC-----GACTCGGCC 3164  
Qy 499 GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPhe 518  
Db 3165 ACGGTCAACAGCTCGACGTCACCTCGCTCGCGCGGATCCGCGCGGAGGCC--- 3221  
Qy 519 ThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGly 538  
Db 3222 ACCGCGCAGACCTGGGTGAT----- 3242  
Qy 539 SerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGly 558  
Db 3243 -----GAACCGCGCGCGCGCGCGGTTCACCGTCCACACCGCGGTGGCGAC 3296  
Qy 559 LysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyThrValAlaAla 578  
Db 3297 GCCCGGTGGACGCTGCACCGCGAG-----GGGTTCCTCGCGCGCGCGGTG---CCC 3347

579 SerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyr 598  
Db 3348 CAGCCCGAAGCGGTGCACACCGCTGGCCCG-----CGGGC-----CGGTGCC 3395  
Qy 599 AlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIle 618  
Db 3396 CGGACCGGCTCGCGGGCGTGGGCGGCGGACGAGTCTTCTGTAAGCCGGAAGTC 3455  
Qy 619 ArgSerAspArgValAsnProLysThrPheTyrAla-----LeuSerAsnGlyThr 635  
Db 3456 GACAGC-----CCTGACGGCTTGTGGGCACACCCCGACCTGTCTGACGCGGTC 3503  
Qy 636 PheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSer 655  
Db 3504 TTCTCCGGCTCGCGCGAGCGG-----AGCGCCAGCGGACCGGATG-CGCGGACCTCGC 3556  
Qy 656 SerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 675  
Db 3557 GGTGACCGCTCGGACGCCACCGTGTGCGCGCTGCTCACCCGCGGACGAGTGTGT 3616  
Qy 676 AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThr 695  
Db 3617 CTGGAGCTCGCGCTTCGACGGTGGCGGATGCGGT-----GCTCACCGC 3664  
Qy 696 GlyValSerSerAla-----ValAsnValGlyPheGlyLysSerAlaProGlySerSer 713  
Db 3665 GGAGTGGTGGAGCTGGCGGAGTGGCTGGCGAGCGGATCCGACGAGTCTGACGCGTCT 3724  
Qy 714 TyrProAlaValPheValValGlyThrIleGlyGlyVal-ThrGlyAlaTyrArgSerAs 733  
Db 3725 GTTTCGGCTTGAGTGGTGGCGGAGCGACCTACGACGCTGCC-----GA 3775  
Qy 733 PAspCysGlyThrTrpValLeuIleAsn-----AspAspGlnHisGlnTyr 749  
Db 3776 CGAGTGGCGGAGGTACACCTCATCCGCCACACACCCCGACGACCCCGACGACCC 3835  
Qy 749 rGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal----- 766  
Db 3836 CACCAACCCCAACACACACCGACCGACCCACACAAACACACACGCGTCTCACCGC 3895  
Qy 767 -----TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle----- 780  
Db 3896 CTCCACACACCACTCATCACCACCAACCACTCATCGTCCACACCAACCGACCC 3955  
Qy 781 ---GlyGlyAlaProSerGly-SerProSerPro-----SerValSerProSerAlas 797  
Db 3956 CCCAGGCGCGCTCACCCTCACCCTCACCCTCACCCTCACCCTCACCCTCACCCT 4015  
Qy 797 erProSerLeuSerPro-----SerProSerProSerSerSerProS 811  
Db 4016 CCACCTCATCGAAACCCACACCCACCCACCCCTCCTCCCTCCTCCTCCTCCTCCT 4075  
Qy 811 erProSerPro-----SerProSerSerSerProSerSerProS 825  
Db 4076 CCACCAACCCCACTTACGCTTACCAACCAACCACTTACCAACCACTTACCACTTAC 4135  
Qy 825 erProSerProSerProSerProSerProSerProSerProSerProSerProS 845  
Db 4136 CACCAACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 4195  
Qy 845 erProSerSerSer-----ProSerProSerSerSerSerSerSerP 860  
Db 4196 CGCATCTCTCATCAGCGGGGCTCGGACCTCTCGCGGATCTCTCGCGGACCTCAA 4255

RESULT 8  
US-09-410-551B-32  
; Sequence 32, Application US/09410551B  
; Patent No. 6503737



```

: GENERAL INFORMATION:
: APPLICANT: KOSAN BIOSCIENCES, Inc.
: APPLICANT: REEVES, CHRISTOPHER
: APPLICANT: CHU, DANIEL
: APPLICANT: CHU, CHAITAN
: APPLICANT: KHOSLA, DANIEL
: APPLICANT: SANTI, DANIEL
: APPLICANT: WU, KAI
: TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
: TITLE OF INVENTION: CONSTRUCTS THEREFOR

```

```

; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 4818

```

Qy	426	TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr	445
Db	3069	-----GACTCGGCGCACCCCGCTTCGCGCAC	3095
Qy	446	GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHis	465
Db	3096	GGAGTCGGCTGCCTCCGGG-----TCGCGGCGCGGGGTGTTTC	3131
Qy	466	IleAlaProMetValLysGlyLeuGluGluThrAla---ValAsnAspLeulleSerPro	484
Db	3132	ACGGGTCCGTGCTGCCCGCGGTGCGGACCGCGGTGTTCATCGCCGAATGGCGGCTCGCC	3191
Qy	485	ProSerGlyAlaProIleuSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAsp	504
Db	3192	GCGCGCGAGCCACC-----GACTGCGCCACCGGTGCAACAGCTCGAC	3233
Qy	505	ValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerVal	524
Db	3234	GTCACTTCGTGCTCCCGCGGATCGCCCGCGGAGGGCC---ACCGCGACAGCTGGGTC	3290
Qy	525	AspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSer	544
Db	3291	GAT-----GACCGCGCGCC	3305
Qy	545	GlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGly	564
Db	3306	GACGGCGCGCGCGCTTACCGTCCACCGCGGTGCGGACCGCGCGCTGCAGCTGCAC	3365
Qy	565	SerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArg	584
Db	3366	GCGGAG-----GGGGTTCCTCCCGCGCGCGCGTG---CCCCGAGCCGAAGCGGTGCAC	3416
Qy	585	PheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsn	604
Db	3417	ACGCGCTGCCCCCG-----CGGGC-----GCGGTGCCCGGAGCGGTGCCCGGG	3464
Qy	605	SerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsn	624
Db	3465	GCGTGGCGAGCGCGGACCGACAGGTCTTCGTGGAAGCGAGTCGACAGC-----	3512
Qy	625	ProLysThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAsp	641
Db	3513	CCTGACGGCTTCGTGGCGACACCCCGACCTGCTCGAGCGGGTCTTCTCCGCGGTGCGGAC	3572
Qy	642	GlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal	661
Db	3573	GGG-----AGCGCGACCGCCAGCGGATG---GCGGACCTCGCGGTGCACGCTCGAGCGC	3625
Qy	662	MetPheHisAlaValProGlyLysGlyGlyAspLeuTrpLeuAlaIleArgSerGlyLeu	681
Db	3626	CACCGTGTGCGCGCTCCCTCACCCCGCGCGACAGTGGTGTGTGAGCTCGCGGCTT	3685
Qy	682	TyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerAla---	700
Db	3686	CGACGTGCGGGAATGCCGT-----GCTCACCGCGAGTCGGTGCAGCTGGG	3733
Qy	701	---ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal	719
Db	3734	CGAGGTGCGTGGCAGCGGATCGGACGAGTCGAGCGTCTCGCTTCGGCTTCAGTGGTT	3793
Qy	720	ValGlyThrIleGlyGlyVal-ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTr	739
Db	3794	GCGGTTGGCGGAGGCCACTACGACGTGTC-----GACGAGCTGCCGAGGCGTA	3844
Qy	739	pValLeulleAsn-----AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAl	755
Db	3845	CACCTTCATCACCGGCCACACACCCCGACACCGCGAGCCCAACCAACCCCAACAC	3904
Qy	755	alleThrGlyAspHisAlaAsnLeuArgVal-----TyrIle	768
Db	3905	ACCCAACGAGCCACACACAAACCAACGCGTCTCCGCGCCCTCAACACCCCTCAT	3964
Qy	768	eGlyThrAsnGlyArgGlyIleValTyrGlyAspIle-----GlyGlyAlaProSe	785

```

Db      3965  CACCACCAACCAACACCTTCATCTGTCACACCAACACGAGCCCCCAGGCGCCGCCGTTCAC 4024
Qy      785  iGly-SerProSerPro-----SerValSerProSerAlaSerProSerLeuSerPro- 802
Db      4025  CGGCTCACCGCAGCGCACAAAACGAACACCCCGCGCATCCCTCATCGAAACCCA 4084
Qy      803  -----SerProSerProSerSerProSerProSerPro----- 814
Db      4085  CCACCCCCACACACCCCACTCCGCCCTCAACCACTCAACACCCCTCAACCAACCCACCTACG 4144
Qy      815  -----SerProSerSerProSerProSerProSerProSerProSerProSerProSerPro 831
Db      4145  CCTCACCAACACACCTCCACACCCGCCCACTCAACCCCATCAACCCACCCACCAACACAC 4204
Qy      831  exProSerProSerProSerProSerProSerProSerProSerProSerProSerProSerSer- 849
Db      4205  CACCACCAACACCCCAACACACCCCTCAACCCCAACCCACCCATCCTCATCACCGG 4264
Qy      850  -----ProSerProSerSerProSerProSerProSerProSerProSerProSerProSerS 866
Db      4265  CGGCTCGGCACCTCGCGGCATCTCGCGCGCACCTCGCGCGCACCTCAACACCCCCACACCTAC 4324
Qy      866  exSerPro 868
Db      4325  CCTCTCCC 4332

RESULT 9
US-09-252-991A-13773/c
; Sequence 13773, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUD
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13773
; LENGTH: 4131
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13773

Alignment Scores:
Pred. No.:      8.98e-06      Length:      4131
Score:          288.50      Matches:      246
Percent Similarity: 31.64%      Conservative: 109
Best Local Similarity: 21.93%      Mismatches: 376
Query Match:      5.62%      Indels:      392
DB:              4          Gaps:      56

US-09-917-376-1 (11-957) x US-09-252-991A-13773 (1-4131)

Qy      37  lleAlaIleThrAlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSer--- 55
Db      4113  GTCAACGTGACCGAGCGAGCGCGCGCGCAATACCAACGCTCCCGCTACCAACGAGCGGTG 4054
Qy      56  -----AsnValAlaIleGlyGlyGlyGlyPheValAspGly 67
Db      4053  GATTCCTCGTGGCTGATCCCGCAGGTGGATCCGAGCAACCGTTCGGTGTATCGCGGC 3994
Qy      68  IleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMet 87
Db      3993  -----ACCGCGAGCGCGCGCAACCC 3973
Qy      88  TyrArgTrpAspAlaAlaAsnGlyAArgTrpIleProLeuLeuAspTrpValGlyTrpAsn 107

```



```
Qy 666 valProGlyValGlyAspLeuTrpLeuAlaSerSerGlyLeuTrpHisSerThr 685
Db 2067 ACACCG-----GTGATCAATGCCAGC 2047
Qy 686 AsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPhe 705
Db 2046 AACCGC-----ACGTGATCACCGCACCGCC-----GAGTTCGGC--- 2011
Qy 706 GlyLeuSerAlaProGlySerSerTrpProAlaValPheValValGlyThrIleGlyGly 725
Db 2010 -----GCCAAGTGTCTCTCCAGCGGCAACGCGCAAC 1978
Qy 726 valThrGlyAlaTyArgSerAspAspCysGlyThrTrpValLeuIleAsnAspAsp 745
Db 1977 CCGATCGCGGAGACCAACCGCGGAGCGGAGTGC---AACTGGACCTTCACCCCGGCGCAGC 1921
Qy 746 GlnHisGlnTyGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg 765
Db 1920 CCGCTGGCCAACGGTACGGTGATCAACCGCGTCCGGAAGACCGCGG----- 1873
Qy 766 ValTyIleGlyThrAsnGlyArgGlyIleValTyArgIleGlyAspIleGlyAlaProSer 785
Db 1872 -----GGCAACGCCAGCGGT---CCGCGC 1852
Qy 786 GlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerProSe 805
Db 1851 AGCACCCAGCGTGACTCGGTGGCGCGTCCGCT---CCGCTGTGTGATCATCAGCGCGCAG 1795
Qy 805 rProSerSerProSerPro-----SerProSerProSe 817
Db 1794 GCGCGCTGTGACCGCGCACCGCGGAGCGGAGCGGAGTGTGCGGTATCGGTCAACGGC 1735
Qy 817 rSerSerProSerSerSerProSerProSerProSerProSerProSerArgSe 837
Db 1734 GACACCGCCAACCCGATCACGTCACGTCGAGCGCGCGGCACTTCACCTTCGCGTTC 1675
Qy 837 r-----ProSerPr 840
Db 1674 GCGCGCGCGTGTATCACCGCGGAACGTATCGCGGGTTCGCGTGGAGCGCGCGGCAAC 1615
Qy 840 oSerAlaSerProSerSerProSerProSerProSerProSerSerProSerSer---Se 859
Db 1614 GTCAGCGCGCGCGCCACCATCAACGCGCGGAGCTTGGCGCGCGCGGACCATCAGCGTCCG 1555
Qy 859 rProSerProTrpProSerSerSerProValSerGlyValValValGlnTyIleLysAs 879
Db 1554 GAAGCGCGCGATACCTGGATCAACGCGCGGAGATCGGGAGCGGC-ATCCAGGTC----- 1501
Qy 879 nAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrG1 899
Db 1500 -----GATGTGACGGTCCGTCGACCATCAGGTCCGCG----- 1468
Qy 899 ySerSerSerValAspLeuSerThrValThrValArgTyTrpPheThrArgAspGlyG1 919
Db 1467 -----CAGGTGTGTCAAGTTC---GCCGGGAGCAACCGGCTA 1430
Qy 919 ySerSerThrLeuValTyAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAl 939
Db 1429 CGAGGCGGAGTGCATACCTCCACCGCGGCGGACATCGCGCGGCAACCTGACCCCT 1370
Qy 939 aSerPhe-----GlySerValAsnPro-----AlaThrProThrAl 951
Db 1369 GACCTGACGCTCCCGGGGCGATGGCCCGTTCCCGGAGGGCGCTCGACCGCTCACCGC 1310
Qy 951 aAsp 952
Db 1309 CGAC 1306
```

## RESULT 10

```
US-09-410-551B-18
; Sequence 18, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
```

```
; APPLICANT: KOSAN BIOSCIENCES, INC.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 4571
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKS synthase fragment
; NAME/KEY: CDS
; LOCATION: (9)..(4559)
; US-09-410-551B-18
```

## Alignment Scores:

```
Pred. No.: 1,01e-05 Length: 4571
Score: 288.50 Matches: 227
Percent Similarity: 34.22% Conservative: 121
Best Local Similarity: 22.32% Mismatches: 390
Query Match: 5.62% Indels: 283
DB: 4 Gaps: 45
```

US-09-917-376-1 (1-957) x US-09-410-551B-18 (1-4571)

```
Qy 9 LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28
Db 1662 CTGACCGAACACGAAGACCGCTGCGCGCTACCTTGGCGGCGTCCCGCGGCGGATATA 1721
Qy 29 AlaAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThr 48
Db 1722 CGGGCTGTGGCATCGACGCTG-----GCGGTGACACGGTCTGGT-GTTCGAGACCGCGC 1774
Qy 49 ThrGlnProTyr----- 52
Db 1775 CGTACTCTTGGAGATGACACCGTCAACCGGACCGCGGTGACCGACCCCGGATCGTGT 1834
Qy 53 -----ThrTrpSerAsnValAlaIleGlyGlyGly-----PheVal 65
Db 1835 TGTCTTCCCGCGAGGGTGGCAGTGGTGGGATGGCAGTGCAGTGCAGTTCGCGATTTCGTC 1894
Qy 66 AspGlyIleValPheAsnGluGly-----AlaProGlyIleLeuTyValArgThrAsp 83
Db 1895 GGTGTGTTCGCCAGCGGATGGCGGAGTGTGCGCGGCGTTCGCGAGTTCGT---GGA 1951
Qy 84 IleGlyGlyMetTyArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
Db 1952 CTGGGATCTGTTCAGGTTCCTGGATGATCCCGCGGTGG-TGGACCGGTTGATG---TGG 2007
Qy 104 ValGlyTrpAsnAsnTrpGlyTyAsnGlyValValSerIleAlaAlaAspProIleAsn 123
Db 2008 TCCAGCCC-GCTTCTCTGGCGATG-----ATGTTTCCCTGGCGCG----- 2048
Qy 124 ThrAsnLysValTrpAlaAlaValGlyMetTyThrAsnSer---TrpAspProAsnAsp 142
Db 2049 -----GTGTGGCAGGCGCGCGGTGTGCGGCGGATGCGGTGATCGCGCATTCGCGAG 2099
```

```
Qy 143 GlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGln----- 156
Db 2100 GGTGAGATCGCGCGCGGTGTGTGGCGGTGCGGTGTCACTACGCGATGCGCGCGGATC 2159
Qy 157 IleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArg 176
Db 2160 GTGACCTTGGCGACGAGCGATCGCCGGGGCTGGCGGGCCGGGGCGGATCGCATCC 2219
Qy 177 LeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 196
Db 2220 GTCGCCCTG---CCCGCGAGAGTGTGAGTGTGTGCGAGGGCC----- 2261
Qy 197 LeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVal 216
Db 2262 ---TGGATCGCGCCCAACAGCGGCCCGCTCCACCGCTGATCGCGGCACCCCGGAAGCG 2318
Qy 217 GlyThrTyrIleAlaAsnProThrAspThrThrGly----- 228
Db 2319 GTCGACCATGTCTCACCGTCTCATAGGACACAGGGGTGCGGGTGGCGGATCACCGTC 2378
Qy 229 ---TyrGlnSerAspIleGlnGlyValValTrpVal-----AlaPheAspLysSer 244
Db 2379 GACTATGCTCGACACCCCGCAGCTGAGCTGATCCGCGAGACTACTCGACATCACT 2438
Qy 245 SerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsn 264
Db 2439 AGCGACAGCAGCTCGACAGCCCGCTCGTGGCGGTGTGTGCGAGCGGCGCACCC--- 2495
Qy 265 ProValPheTrp-----SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAla 282
Db 2496 -----TGGGTGCGACAGCCCGCTGGAGCGGGAGTACTGTACCGGAACCTCGCGTAA 2546
Qy 283 ProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIle 302
Db 2547 CCGGTGCTTTCACCCCGCGCTGAGCGGTCAGCGGTCAGGCGCCAGGCGGACACCGTTCGTC 2606
Qy 303 AlaThrSerAsnThrGlyGlyProTyrAspLysSerSerGlyAspValTrpLysPheSer 322
Db 2607 GAGTCAAGCCAGCCCGGTGTGTCAGGCGATGAGCAGCATGTCTGTCAGGTGCC 2666
Qy 323 Val-----ThrSerGlyThrTrpTrpThrArgLysSerProValProSerThrAspThr 339
Db 2667 ACGTGGTGTGTGACGAGCGCGACCGCCCGGATG-----CTCACCGCCCTG 2714
Qy 340 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 359
Db 2715 GCACAGCCCTAT-----GTCCAGCGGTGTACCGTCGAC----- 2747
Qy 360 MetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly 379
Db 2748 -----TGGCCCGCCATCTCTC----- 2762
Qy 380 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 399
Db 2763 GGCACCAACCAACCCGGGTACTGGACCTTCCGACCTACGCTTCCAAACACAGCGGTAC 2822
Qy 400 ValLeuAspLysSerAlaGluProTrpTrpLeuThrPheGlyValGlnProAsnProVal 419
Db 2822 ----- 2822
Qy 420 ProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 439
Db 2823 -----TGGCTCGAGTCGCGACGC-----CCGGCGGATCCGAC 2855
Qy 440 -----ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys 457
Db 2856 GCGGGCCACCCCGTGGGCTCGGTATCGCCCTCGCCCGG----- 2897
Qy 458 TrpAspSerGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAla 477
Db 2898 -----TCGCGGCGCGGGTGTTCACGGGT----- 2921
Qy 478 ValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 497
```

---

```
Db 2922 -----TCCGTGCCGACCGGTGTCGACCGCGCGGTGTGTGTCGCCGAGCTG 2966
Qy 498 GlyGlyPheThrHisAlaAspVal-ThrAlaValProSerThrIlePheThrSerProVal 517
Db 2967 GCGCTGCGCGCGCGGCGCGGTGCGACTGCGCCACCGTTCGAGCGGTCCACATCGCTCC 3026
Qy 517 lPheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAl 537
Db 3027 GTGCCCGCGCGCGGCGCATGGCGGACGACCGTACAGACCT----- 3069
Qy 537 aGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 557
Db 3070 -GGGTCCGAC-GAGCGCGCGGACGACGCGCGCGGTTCACCGTGCACACCGCGCACCG 3127
Qy 557 yGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyValAla 577
Db 3128 CGACGCGCGGTGGAGCGTGCACGCGGAG---GGGTGCTGCGCGCCCGTGGCAGC--- 3179
Qy 577 aAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValVa 597
Db 3180 -GCCTGCCGATGCGCGCGGACGCGAGTGGCCCCA-----CGGGC-----GCGGT 3226
Qy 597 lTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGly 617
Db 3227 GCCCGCGACGCGGTGCGCGGTGTGTGCGCGCGGGGACGAGGTCTTCGCCGAGCGCA 3286
Qy 617 nIleArgSerAspArg-----ValAsnPro-----LysThrPheTyrAl 630
Db 3287 GGTGCGGACGCGGACGCGGTTCGTGTGTCACCCCGACCTGTCGACGCGGTCTTCTCCG 3346
Qy 630 aLeuSerAsnGlyThr----- 635
Db 3347 GGTGCGGACGCGAAGCCGCGGCGCGGATGGCGGACCTGACGCGTGCACGCGTCGGA 3406
Qy 636 -----PheTyrArgSerThrAspGlyGlyValThrPheGlnPr 648
Db 3407 CGCCACCGTACTGCGCGCTGCTCACC CGCGCGCACCGAGCATGGATTTCGCCG 3466
Qy 648 oVal---AlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValPr 667
Db 3467 CTTCGAGCGGCGCGCGCTGCGGTACTCACC CGCGGAGCGGTGACGCTCGCGGAGGTGC 3526
Qy 667 oGlyLysGluGlyAsp-----LeuTrpLeuAlaAla 678
Db 3527 GTCACCGTCCGCTCCGAGGAGTGGACGCGCTGACCGGTGGAGTGGCTCGCGTCCG 3586
Qy 678 rSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValse 698
Db 3587 CGAGCGGTCTAC-----GACGGTGACCTGCCCGAGGACATGTC-----CTGAT 3631
Qy 698 rSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPh 718
Db 3632 CACCGCGCGCACCCCGACGACCGGAGACATACCCACCGCGCGCCACACCGCGCGCAC 3691
Qy 718 eValVal----- 720
Db 3692 CCGGTCTCTGACCGCTGCAACACCACTCACCACCGACGACACACCTCATCTGTCGA 3751
Qy 721 -----GlyThrIleGlyGlyValThrGlyAlaTyrArgSerAs 733
Db 3752 CACCACACCGACCCCGCGCGCATCGTACCGGCTCACCACCGCGCCACGCAACGA 3811
Qy 733 pAspCysGlyThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGly 753
Db 3812 ACACCCCGACCGCATCGCGCTCATCGAAACCGACCCACCCCGCGCGCCACCGCGCGC 3871
Qy 753 yGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyr----- 767
Db 3872 CCAACTCGCACCGCTCGACCAACCCCGCTCCCTCACCACCGACACCCCTCCACCGCC 3931
Qy 768 -----IleGlyThrAsnGlyArgGlyIleValTyr-GlyAspIleGlyGly 783
```



ORGANISM: Mycobacterium tuberculosis  
 ; OTHER INFORMATION: H37Rv  
 US-09-103-840A-1

## Alignment Scores:

Pred. No.: 0.0493 Length: 4411529  
 Score: 282.50 Matches: 278  
 Percent Similarity: 32.66% Conservative: 111  
 Best Local Similarity: 23.34% Mismatches: 377  
 Query Match: 5.50% Indels: 432  
 DB: 3 Gaps: 59

US-09-917-376-1 (1-957) x US-09-103-840A-1 (1-4411529)

Qy 4 SerGluAsnIleArgLeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThr 23  
 Db 372877 TCCAGCGACTCCAGGCTTCGGCAACTCGACCGGGGACTGTC----- 372836  
 Qy 24 AlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerPro 43  
 Db 372835 -----GTCGGGCTTCTTCA-ATAGCGGCTGCGCAGCGCT 372801  
 Qy 44 AlaHisAlaAlaThrThr--GlnProTyrr-----ThrTrpSerAsnValAla 58  
 Db 372800 CGGGCTTCGGCAACTTCGGGGCGCGCGCTGGCTTTATCAACCTGGTCTCC-ACAAAG 372742  
 Qy 59 IleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAla-----ProGly 76  
 Db 372741 TCGGGAATGTCGGCTTCCTC-----AACGTGGCGGCTGGGATCGGGT 372697  
 Qy 77 IleLeuTyrrValArgThrAspIleGlyMetTyrrArg- TrpAspAlaAlaAsnGlyAr 96  
 Db 372696 GTGGCGAATGTGGCAACACCACTCGGGTATCTACAACTGGCGCAGCTCGGACTCTCG 372637  
 Qy 96 gTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrrAsnGlyValValse 116  
 Db 372636 ACGCCGCGCTTAACCTCG-----GGTTGGCAA----- 372609  
 Qy 116 rIleAlaAlaAspProIle-----AsnThrAsnLysVal----- 127  
 Db 372608 -----ATATCGGAACCAATATTCGGCGCTGTCGGCGACGGCGGTACTCGGGCTATT 372553  
 Qy 128 -----TrpAlaAlaValGlyMetTyrrAsnSerTrpAspProAsnAspGlyAlaIleLe 146  
 Db 372552 AACTTGGGCTTGCCCAACACCGCAACTCACTGCGGCTTCGCAAGTCTCGCGGGCTTT 372493  
 Qy 146 uArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGI 166  
 Db 372492 AACTTCG----- 372486  
 Qy 166 yAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnI 186  
 Db 372485 -----GCGGCGCCACCACTCGGCCCAACCAAC 372460  
 Qy 186 eLeuTyrrPheGlyAlaPro--SerGlyLysGlyLeuTrpArgSer----- 200  
 Db 372459 GTGGCATCGGAACACCGGAATCTTCGATGTGGGCTGCGGACCTCGGCAGCTACAAAC 372400  
 Qy 201 -----ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVa 216  
 Db 372399 ATCGGCTTCGGAACCTTCGGCAGCAGCACTTCGGCTTCGGCAACTTCG-----GC 372349  
 Qy 216 lGlyThrTyrrIleAlaAsnProThrAspThrThrGlyTyrrGlnSerAspIleGlnGlyVa 236  
 Db 372348 AGCTACAACTACGCTTCGGCAACGTCGCAACG-----AC 372313  
 Qy 236 lValTrpValAlaPheAspLysSerSer----- 245  
 Db 372312 AATCTGGGTTTCCTCAACCGGGCGGGCGCAACATCGGCTTTGGAAACACCGGAGCAAC 372253  
 Qy 246 -----SerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAs 264  
 Db 372252 AATGTCGGCTTTGGGAACACCGGCGAGCAACATGTCGGCATCGGGCTCAGCGGCACCGGA 372193

Qy 264 nProValPheTrpSerArgAspGlyGlyAla----- 274  
 Db 372192 C-----AGATCGGGTTCGGCAGCTTCAACTCGGGCAGCGGAACATCGGC 372148  
 Qy 275 -----ThrTrpGlnAlaValProGlyAlaProThrGlyPh 286  
 Db 372147 CTGTTCAACTCGGCAGCAACCAATCGGATTCCTCAATCCGGCAGCGCACTTCGGC 372088  
 Qy 286 eIleProHisLysGlyValPheAspProValAsnHisValLeuTyrrIleAlaThrSerAs 306  
 Db 372087 -ATCGCAACTCGGCAGCTTCAAC-----ACTGGCATCGGAAA 372050  
 Qy 306 nThrGlyGlyProTyrrAspGly-----SerSerGlyAspValTrpLysPheSerValTh 324  
 Db 372049 CACCGCAACCAATACCGGCTTATCACTCCGGCGAC-----GTCAA 372005  
 Qy 324 rSerGlyThrTrp- ThrArgIle-----SerProValProSerThrAspThrAlaAsnA 342  
 Db 372004 CACGGCGCCTTCAACCGGGCAGCTTCAACACCGGCTAGCTTCAACCGCAGCTTCAA 371945  
 Qy 342 sPTyrPheGlyTyrrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValA 362  
 Db 371944 CACCGGTGGCT---TCAATCCGGGCAATACCAACACCGGCTACCTCAACTTGGCAACTA 371888  
 Qy 362 laThrGlnIleSerTrpTrpProAspThrIleIle-----PheArgSerThrAspGlyG 380  
 Db 371887 CAACACCGGCTCGCCACACACCGGGCAGCTTGACACCGGGGCTTTCATCACCGGAAACTA 371828  
 Qy 380 lyAlaThrTrp-----ThrArgIleTrpAspTrpTrpSerTyrrProA 394  
 Db 371827 CAGCAACGGGTGTTCCTTAAGCGGCGATTACACAGGCGCTGG----- 371787  
 Qy 394 snArgSerLeuArgTyrrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValG 414  
 Db 371786 -----TCGGCCTCAACCTGGTGATCATGATGCTCT 371756  
 Qy 414 lnPro-----AsnProProValProSerP 422  
 Db 371755 CCCATAAGCTTCGGGTGAATATTCCTCATCGATATCCGATCACCGCTCGGCCGCA 371696  
 Qy 422 roLysLeuGlyTrp----- 426  
 Db 371695 CATCACCTTATGGCGTCAAGATTCCGCCACCGCGATATCGTCTTTCGTCAATAGC 371636  
 Qy 427 -----MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrr--GI 444  
 Db 371635 GGGCCAGCGAGCCCACTTTGGCCCCATTAACCATTCGGAACATCACGCTTGTGGCCCCAC 371576  
 Qy 444 yThrGlyAlaThrLeuTyrrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlnI 464  
 Db 371575 GACGACAGTCCCATAGAGGGCGGATACCGCGATACCATTAACGTGGCGGTGGGCCCAT 371516  
 Qy 464 eHisIle-----AlaProMetValLysGlyLeuGluThrAlaValAsn-- 479  
 Db 371515 TAGGATCCGCTCATCATGATCCCGCGGCGCAGTTTCGGAACCTCGACCAACCC 371456  
 Qy 479 ----- 479  
 Db 371455 GTCGTCAAGTTTCTTAATACCGCGCGCGCTCGGGCTTCGGCAACTTCGGCGG 371396  
 Qy 480 -----AspLeuIleSerProProSerGlyAlaProLeuIleSerAl 493  
 Db 371395 CGCAATTCGGGCTTTTGGAACTTGGCTTCGGCGACCTTCGGGGCGTTCCTCAA 371336  
 Qy 493 aLeuGlyAspLeuGly-----GlyPheThrHi 502  
 Db 371335 CGTCGGCGCTTGGATCAGCTTGGCGCAACGTGGGCAACCGTCTCGGGGTTCACAA 371276  
 Qy 502 sAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyTh 522  
 Db 371275 CACCAGCAGCTCGGACCTCGCGAGCGCGGCTTCAATTCAGGCGCTGGCCCAACATCAGCAC 371216



```
QY 522 rSerVal----- 524
Db 371215 CAGTATCGCGGCTGTGCGCAGACGACGCGGCACCATGCTCTCAACCTGGGCTTGGC 371156
QY 525 -AspTyrAlaGluLeuAsnProSerIleIleValAlaGlySerPheAspProSerSe 544
Db 371155 AAACACCGGCACCTCAACCTCGGCATT-----GC 371126
QY 544 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp----- 561
Db 371125 AAACCTCGGCGACTACACATCGCTTTCGAACCTCGGCGAGCCCACTTCGCGACGCG 371066
QY 562 ----PheGlnGlySerGluProGlyValThr----- 572
Db 371065 CAATATCGGTGGCAACACATCGGCGCGCAACACCGGAATATTCGACATCGTTTGGC 371006
QY 573 ----GlyGly-ThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProG 590
Db 371005 AAATCGGGCAGTTACACATCGCTTCGGAACCTTCGCGGATGACAACTGGGCTTC-G 370947
QY 590 LysAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerG 610
Db 370946 GAACCTCGGCAGC-----TACACGTGGCTTCGGAACCTTGGGCAACGACCAACC 370896
QY 610 LnglyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrA 630
Db 370895 TGGGCTTC--GCCAACACCGGC----- 370876
QY 630 laLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProVala 650
Db 370875 ----AGCAACAATATCGGCTTCGCGAACACCGGCAGC-----AACAAATATCG 370833
QY 650 laAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysG 670
Db 370832 GCATTGGCTTCAGGGGCGCGCAGATCGG-----TTGGCTCCTCGAATCTCGCA 370779
QY 670 luGlyAspLeuTrpLeu-----AlaAlaSerSerGlyLeuTyrHisSerThrA 686
Db 370778 GCGGAAACATCGGCTTGTTCACCTCGGCGCAGCGGAACATCGGCTTTTCAACTCGGCA 370719
QY 686 snGly-----GlySerSerTrps 692
Db 370718 ACGGAAACGTTGGCATCGGCAACACCGGCACCGCAACTTCGGGCTTGGAAACACCGGCA 370659
QY 692 erAlaIleThrGlyVal-----SerSerAlaValAsnValGlyPheGlyLysSerAla- 709
Db 370658 GCACCAACACCGCTTCTCACTCGGCGACGTCAATACGGTATCGGCAACACCGGCA 370599
QY 710 -----ProGlySerSerTyrProAlaValPheValValGlyT 722
Db 370598 GCTTCAACACCGCGAGCTTCAATCGGCGCATTCACACACCGGGGATTTCACACCGGCA 370539
QY 722 hrIleGlyValThrGlyValaTyrArgSerAspAspCysGlyThrTrpValleuI 742
Db 370538 GCTACAAAC-----ACGGGACTCGGAACACCGCGGATGTTCACAC----- 370498
QY 742 leAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaA 762
Db 370497 -----GGCGCTTCATCTCGGCGAGCTACAGCA 370470
QY 762 snLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyG 782
Db 370469 AC-----GGGTTCTTGTGGAGTGAATATTCAGGGGCTCATCGCT 370428
QY 782 lyaAlaProSerGlySerProSer-----ProSerValSerProSerAlaSerProS 799
Db 370427 T-GCAGCGCGCGCTAGCGATTTCGGAATCGCCCTAACCTTTG-GCGTCGACATCCCGAT 370370
QY 799 erLeuSerProSerProSer-----ProSerSerPro----- 810
Db 370369 ACATATACCCATCAACATCGACCGCGGGTTCGTCACCCCTCCAGCGGCTTCAGCATCGTAGC 370310
QY 811 -----SerProSerProSerProSerSerProSerSerProSerP 826
```

```
Db 370309 TGCGGAAATAATATCGACTTCAACCCCATCATCCCGACCATATATCACCTTGGC 370250
QY 826 ro-----SerProSerProSerProSerProSerArgSerProSerProSerAlaAs 843
Db 370249 CACGGCGCGCATCACCGTGGCGGACCCA-----CCACCTCGATCGGTATCACCG----- 370200
QY 843 erProSerProSerProSerProSerProSerProSerProSerProSerP 860
Db 370199 --CCAGCGCGGTATCGGCTCCATCACCATCCGATCATCGACATTCGCGGACATCGGG 370142
QY 860 roSerProThrProSerProSerProValSerGlyValLysValGlnTyrLysAsnA 880
Db 370141 CTTGGCAACTCG-ACCACCTAGTCGTCGCGGC-----TTCTTCAACT 370098
QY 880 sn-----AppSerAlaProGlyAspAsnGlnIleLysProGlyLeuGln----- 894
Db 370097 CCGAGCGGCGCAGCGCTGCGGCTTTTGAAGTGGTTCGCGCGCCTCAGGGATTCGG 370038
QY 895 -----ValValAsnThrGly----- 899
Db 370037 GTTATCTCAATGTCGTCGCTGGGTGGGTGTGACTAAACGTGGGTACACCGCTCTCGG 369978
QY 900 -----SerSerSerValAspLeuSerThrValThrValArgTyrTrpPheThrA 916
Db 369977 GTTTCTACAACGCGACGCGCTTGGACCTCGTGCACCGCGCTTTCCTCCCGCTCTCATGC 369918
QY 916 rGAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaIleGlyCysGlyA 936
Db 369917 GCGHAC-----GGTATGGGACGAGTACTGTGAACCTTGGCTGGCAAACTGGGCGACGAATA 369861
QY 936 snIleArgAlaSerPheGlySer 943
Db 369860 AC-----GCCGGCTTCGGCAAC 369844

RESULT 14
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred No.: 0.0552 Length: 4403765
Score: 281.50 Matches: 278
Percent Similarity: 32.66% Conservative: 111
Best Local Similarity: 23.34% Mismatches: 377
Query Match: 5.48% Indels: 432
DB: 3 Gaps: 59

US-09-917-376-1 (1-957) x US-09-103-840A-2 (1-4403765)
QY 4 SerGluAsnIleArgLeuThrMetArgSerArgLeuValSerLeuAlaIleThr 23
```

```
Db 372934 TCACGCGACTCCAGGCTTCGGCAACTCGACCGCGGACTGCTC----- 372893
Qy 24 AlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerPro 43
Db 372892 -----GTCGGGCTTCTTCA-ATAGCGGCGCTGGCAGCGCCT 372858
Qy 44 AlaHiAlaAlaThrThr-----GlnProTyr-----ThrTrpSerAsnValAla 58
Db 372857 CGGGCTTCGGCAACTTCGGGGGCGCGCGCTCGCGCTTTATGAACCTGGTCTCC-ACAACG 372799
Qy 59 IleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAla-----ProGly 76
Db 372798 TCGGGAATTCGGGCTTCCTC-----AACCTCGCGCGCTGGGATCGGGT 372754
Qy 77 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArg-TipAspAlaAlaAsnGlyAr 96
Db 372753 GTGGGAGTGGGGCAACACCATCTCGGTATCTACACGTGGCAGCTCGGACCTCTCG 372694
Qy 96 GTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSe 116
Db 372693 ACGCCGCGCTTAACCTCG-----GGTTGGCAA----- 372666
Qy 116 rIleAlaAlaAspProfile-----AsnThrAsnLysVal----- 127
Db 372665 -----ATATCGGAACCAATATTCGCGGCTGCTCGCGCAGCGCGGTACTCGGGCTATT 372610
Qy 128 -----TrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLe 146
Db 372609 AACTGGGCTGGCCMACCGGCAACCTCAACGTGGGCTTCGCAAGTCTCGCGCGCTTT 372550
Qy 146 uArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGI 166
Db 372549 AACTTCG----- 372543
Qy 166 yAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnI 186
Db 372542 -----GCGGGCCCAACCATCGGCCCAACAAC 372517
Qy 186 eLeuTyrPheGlyAlaPro---SerGlyLysGlyLeuTrpArgSer----- 200
Db 372516 GTCGGCATCGGAACACCGGAATCTTCGATGTCGGCTCGCGAACCCTGGGCACTTCG- 372457
Qy 201 -----ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVa 216
Db 372456 ATCGGCTTCGGAACCTTCGGCAGCAACAACCTGGCTTCGGCAACTTCG-----GC 372406
Qy 216 lGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVa 236
Db 372405 AGCTACAACATCGGCTTCGGCAACGTCGGCAACG-----AC 372370
Qy 236 lValTrpValAlaPheAspLysSerSer----- 245
Db 372369 AATCTGGGTTTCGTAACCGGGGCGCGCAACATCGGCTTTCGGAACACCGGCAACAAC 372310
Qy 246 -----SerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAs 264
Db 372309 AATGTCGGCTTTGGGAACACGGGCAACAATGTCGGCATCGGGCTCACGGGCAACCGGA 372250
Qy 264 nProValPheTrpSerArgAspGlyGlyAla----- 274
Db 372249 C-----AGATCGGTTTCGGCAGCTTCAACTCGGGCAGCGGAACATCGGC 372205
Qy 275 -----ThrTrpGlnAlaValProGlyAlaProThrGlyPh 286
Db 372204 CTGTTCACTCGGGCAGCAACAACATCGGATCTTCAATTCGGCAGCGCAACTTCGGC 372145
Qy 286 eIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAs 306
Db 372144 -ATCGAATCTCGGGCAGCTTCAAC-----ACTGGCATCGGNA 372107
Qy 306 nThrGlyGlyProTyrAspGly-----SerSerGlyAspValTrpLysPheSerValTh 324
Db 372106 CACCGCAACCAATACCGGCTTATCAACTTCGGGCGAC-----GTCAA 372062
```

QY 573 -----GlyGly-ThrValAlaAlaSerAlaAspGlySerArgPheValTyrAlaProG 590  
Db 371062 AAATCTGGCAGTACCAATCGCTTCGGAATCTTCGGATGACAACTCGGCTTC-G 371004  
QY 590 lyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTyrAlaAlaSerG 610  
Db 371003 GAACCTCGGCAGC-----TACACGTCTGGCTTCGGAATCTTCGGCAACGACACACC 370953  
QY 610 lnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLoLysPheTyrA 630  
Db 370952 TGGGCTTC--GCCAACACCGGC----- 370933  
QY 630 laLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValA 650  
Db 370932 -----AGCAACAATATCGGGTTCGCAACACCGGCAGC-----AACAAATATCG 370890  
QY 650 laAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysG 670  
Db 370889 GCATTGGGCTCAGCGGGCAGCGGCAGATCGG-----TTCGGCTCCCTGAATCTGGCA 370836  
QY 670 luGlyAspLeuTyrLeu-----AlaAlaSerSerGlyLeuTyrHisSerThrA 686  
Db 370835 GCGGAACATCGGCTTGTTCACACTCGGGCAGCGGAACATCGGCTTTTTCACACTCGGCA 370776  
QY 686 snGly-----GlySerSerTyrPs 692  
Db 370775 ACGGAACGTTGGCATCGGCACACACCGGCACCGCAAACTTCGGGCTTGGAAACACCGGCA 370716  
QY 692 erAlaIleThrGlyVal-----SerSerAlaValAsnValGlyPheGlyLysSerAla- 709  
Db 370715 GCACCAACACCGGCTTCTTCACTCGGCGACGTCAATACCGGTATCGGCAACACCGGCA 370656  
QY 710 -----ProGlySerSerTyrProAlaValPheValValGlyT 722  
Db 370655 GCTTCAACACCGCAGCTTCAATCGGGCGATTCACACCGGGATTCACCCAGGCA 370596  
QY 722 hrIleGlyGlyValThrGlyValTyrArgSerAspCysGlyThrTyrValLeuI 742  
Db 370595 GCTACAAC-----ACGGGACTCGGAACACACCGCGCATGTGACACC----- 370555  
QY 742 leAsnAspAspGlnHisGlnTyrGlyAsnTyrGlyGlnAlaIleThrGlyAspHisAlaA 762  
Db 370554 -----GCGCGCTTCATCTCCGGCAGCTACAGCA 370527  
QY 762 snLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyG 782  
Db 370526 AC-----GGGTCTTGTGGAGTGGAAATATACAGGGCTCATTTGGCT 370485  
QY 782 lyAlaProSerGlySerProSer-----ProSerValSerProSerAlaSerProS 799  
Db 370484 T-GCAGCGCGGCTAGCGATTCCGGAATCGCCCTAACCTTTG-GCGTCGACATCCGAT 370427  
QY 799 erLeuSerProSerProSer-----ProSerSerSerPro----- 810  
Db 370426 ACATATACCATCAATCAATCGACCGGGGTGTCACCTCCAGGGCTTCAGCATCGTAGC 370367  
QY 811 -----SerProSerProSerProSerProSerProSerProSerProSerP 826  
Db 370366 TGCCGAAATAATATGACTTCACTCCCATCATCATCATCATCATCATCATCATCATCATCAT 370307  
QY 826 ro-----SerProSerProSerProSerProSerProSerProSerProSerAlas 843  
Db 370306 CACGGCGCGCATCACCGTGGCGGACCCA---CCACCTCGATCGGTATCACCG----- 370257  
QY 843 erProSerProSerProSerProSerProSerProSer-----ProSerSerSerP 860  
Db 370256 --CCAGCGCGGTATCGGCTCCATCATCATCATCATCATCATCATCATCATCATCATCAT 370199  
QY 860 roSerProThrProSerSerSerProValSerGlyGlyValLysValGlnTyrLysAsnA 880  
Db 370198 CTTTGGCAACTCG-ACCACATAGTCGTCGCGGC-----TTCTTCAACT 370155  
QY 880 sn-----AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGln----- 894

Db 370154 CCGAGCGGCAGCGCGCTCGGCTTTTGAACGTGGTCCGCGGCTCAGGATTCGG 370095  
QY 895 -----ValValAsnThrGly----- 899  
Db 370094 GTTATCTCAATGTGGTGGCTGGGTGGGTGAGTAACTGAGTGGTACACCGTCTCGG 370035  
QY 900 -----SerSerSerValAspLeuSerThrValThrValArgTyrTyrPheThrA 916  
Db 370034 GTTCTTCAACGCGGCGGTGACCTCTGTGACGCGCGGCTTTTCCTCCGCTCTCATGC 369975  
QY 916 rAspGlyGlySerSerThrLeuValTyrAsnCysAspTyrAlaAlaIleGlyCysGlyA 936  
Db 369974 GCGAC-----GGTATGGGACCATGACTCTGAACCTTGGGCTGCAAACTGGGCAACA 369918  
QY 936 snIleArgAlaSerPheGlySer 943  
Db 369917 AC-----GCGGCTTCGGCAAC 369901  
RESULT 15  
US-09-252-991A-13873/c  
; Sequence 13873, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13873  
; LENGTH: 3129  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13873  
Alignment Scores:  
Pred. No.: 1,73e-05 Length: 3129  
Score: 280.00 Matches: 245  
Percent Similarity: 33.12% Conservative: 114  
Best Local Similarity: 22.60% Mismatches: 396  
Query Match: 5.45% Indels: 233  
DB: 4 Gaps: 61  
US-09-917-376-1 (1-957) x US-09-252-991A-13873 (1-3129)  
QY 36 ProfileAla-----IleThrAlaSerProAlaHisAlaAlaThrThr 49  
Db 3113 CCGATCGCCAAACGGTACTGTGTCAACGTGTGGCGGAGGAGCGCCGCGTTAACAGCAGC 3054  
QY 50 GlnProTyrThrTyrPsrAsnValAlaIleGlyGlyGlyPheValAspGlyIleVal 69  
Db 3053 CCGCGCGGCGAG-----GTGACC 3036  
QY 70 PheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArg 89  
Db 3035 GTCGATTCCAGCGCGCGCGCGCGGTGATCAACCGAGCAACGCGGTCTCATCAGC 2976  
QY 90 TrpAspAlaAlaAsnGlyArgTyrIleProLeuLeuAspTyrValGlyTyrAsn----- 107  
Db 2975 GGCACCGCGGAGCGGTGCGACCGTACCGGAT---GCCGCGGCGCAACCGGATA 2919  
QY 108 -----AsnTyrGlyTyr----- 111  
Db 2918 GGGCAGGTACCGCGGAGCGGCGGCACTGGAGCTTCACGCCGGGCAACCGCGCGGCC 2859  
QY 112 AsnGly-----ValValSerIleAlaAlaAspProfile---AsnThrAsnLysValTyr 128

2858 AACGGCAGGTGATCTGCGCACCGGCACCGACCGGCAATACCGGCCCGCAGGCC 2799  
Qy  
129 Ala-----AlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAla 144  
Db  
2798 GCCACCAAGGTGGACGGGTGGCGCCGCGCGGTGATCGATCGGACGACGCGCACG 2739  
Qy  
145 IleLeu----- 146  
Db  
2738 ACCATCAGCGGCACCGGAGCGCGCGGCGCAAGGTGATCTCTACCGACGCGCAACG 2679  
Qy  
147 -----ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro---LeuPro 161  
Db  
2678 CCGATCGCGGAAACCAACCGCGGACGCGGACGCGGCACTGGACCTTCACGCGCGCACGCG 2619  
Qy  
162 PheLysLeuGly-----GlyAsnMetProGly 170  
Db  
2618 CTGGCCCAACGGCAGCGGTGTCACCGCGTGGCCCGCAGGACCTGTGGGGCAATACCGGCCG 2559  
Qy  
171 ArgGlyMetGlyGluArgLeuAlaValAspProAsn-----Asn 183  
Db  
2558 CAGGGCAGCACTACCGTGGACGGGTGGCGCGGAACACGCGTGTGTGTCATCCGAGCAAC 2499  
Qy  
184 AsnAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeu----- 197  
Db  
2498 GGCACACCTGCTCAACGGTACCGCGGACCGCGGACGACCGTGTGACCTTGACCGACGCAAC 2439  
Qy  
198 -----TrpArgSerThrAspSerGlyAlaThrTrp----- 207  
Db  
2438 GGCACACCGCATCGGCAGACCGCGCGATGGCGAGCGGCACTGGAGCTTCACGCGCGCG 2379  
Qy  
208 SerGlnMetThrAsnPheProAspValGlyThrTyrIle---AlaAsnProThrAspThr 226  
Db  
2378 TCGCAACTACCCAAC-----GGCACCGTGTGTCAACGTGACCGCGGACGCGCC 2331  
Qy  
227 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 246  
Db  
2330 GCCGGCAATACCAAC-----GCTCCGCGTACCAACGCGGATGCTCTCGTGGCG 2280  
Qy  
247 SerLeuGlyGln-----AlaSerLysThrIlePheValGlyValAlaAspProAsn 263  
Db  
2279 TCGATCCGCGAGGTGGATCCGAGCAACGGTTCGCTGATCAGCGCACCGCGGACGCGCG 2220  
Qy  
264 AsnProValPhe-----TrpSerArgAspGly 272  
Db  
2219 AACACCATCATCATCCCGATGGCAACCGGACCGGATGGCCAGGTCAACGCGACGCG 2160  
Qy  
273 GlyAlaThrTrpGlnAlaValProGlyAlaPro----- 283  
Db  
2159 AGCGTAACTGGTCTTCACTCCAGGATCCCGTGGCGGATGGCAGCGGTGTCAACGCGT 2100  
Qy  
284 -----ThrGlyPhe 286  
Db  
2099 GTGGCGCGCACCCCAAGCAATGTCGACAGTGCAGCGCGCGGTGATCACTGTGGATGGCGTG 2040  
Qy  
287 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 306  
Db  
2039 GCCCGCGCGCGCGGTGATGATCCGAGCAACGCGCACCGGACGAGATGACGGGTACCGCGGAG 1980  
Qy  
307 ThrGlyGlyProTyr-----AspGlySerSerGlyAspValTrpLysPheSerVal 323  
Db  
1979 GCCGGCGCAGCGGTGATCTCTACCGATGGCGGGGCAACCGATCGCGCCAGCCACCGCC 1920  
Qy  
324 Thr---SerGlyThrTrpThrArgIleSerProValPro----- 335  
Db  
1919 GACGGCAGCGGCAACTGAGCGTTTCACCGCGGCAACCGCGTGGCCAAACGCGCATCGGTGATC 1860  
Qy  
336 -----SerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAsp 352  
Db  
1859 AACCGCGCGCGGACCGCGCGGCGCAATACGACGCGGTCCGCGCACGCGTCACTGAT 1800  
Qy  
353 ArgGlnHisPro-----AsnThrIleMetValAlaThrGln 364  
Db  
1799 GCCATCGCGCGCGCGCGGTGATCAATCCGAGCAACGAGGTGCTGTCATCAGCGGTACG 1740

365 IleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly----- 380  
Db  
1739 CGCGAAGCGCGGCGCACCGGTGATCTCT-----ACCGACGCGCAACGCGCAACCGCATC 1689  
Qy  
381 -----AlaThrTrpThrArgIleTrpAspTrpThr---SerTyrProAsnArg----- 395  
Db  
1688 GCCCAGGTACCGCGCGCAGCGGCAAGTGGGCTTTACCGCGCGCACCGCGTGGCGC 1629  
Qy  
396 -----SerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPh 411  
Db  
1628 AATGGCAGCGGTGATCAATGCGCTGG-----CCGAGGACGCGCGCGGCAACAC 1581  
Qy  
411 eGlyValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAlaMe 431  
Db  
1580 AGCAGTC-----CCACCAGCGCCA---CCGTCGACTCGCTGGCGGCCAGCA-GCCCC 1534  
Qy  
431 tAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr-----GlyAlaTh 448  
Db  
1533 GGTGATCGATCCGAGCAACGCGTAGC---GTGATCGCGGTACCGCGGAGGCTGGTGGCCAC 1477  
Qy  
448 rLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaPr 468  
Db  
1476 GGTGATCTCTACC-----GACGGCAACGCGCAAC-----CC 1447  
Qy  
468 oMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSerGlyAl 488  
Db  
1446 GATCGCGCAGTCAACCGCGGATGGCAGCGGCACTGGAGTTCAGCGCC-----GGCAC 1393  
Qy  
488 aProLeu-----IleSerAlaLeuGly---AspLeuGlyGlyPheThrHi 502  
Db  
1392 GCCGCTGTCCAATGGCAGCGGTGTCATCGGTGGCCCGGACGCTGCGGCAACACACAG 1333  
Qy  
502 sAlaAspVal---ThrAlaValProSerThrIlePheThrSerProValPhe-----Th 519  
Db  
1332 CGGCGCGCGCACGACGCGTGGACTCGGTGGCGCGCGCGCGGTGATCGACCCCGAG 1273  
Qy  
519 rThrGlyThrSerVal---AspTyrAlaGluLeuAsnProSerIleIleValArgAlaGl 538  
Db  
1272 CAACGCGCAGCTGATCGCGGTGATCGCGAGACCGCGTGGCGGTGATCTCACCAGTGG 1213  
Qy  
538 ySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 558  
Db  
1212 CGGCGGCAACCGATCGCGCAGGCC-----ACCGCGATGGCGAG 1174  
Qy  
558 yLysAsnTrp-----PheGl 563  
Db  
1173 CGGCAACTGGAGCTTCAACCGCGGCAACCGGTGGCCAAACGCGACGCGTGAATCAATGCGGT 1114  
Qy  
563 nGlySerGluProGlyValThrGlyThrGlyGlyThr----- 575  
Db  
1113 GGCCAGGATCCGCGCGCAATACCGCGCGCGCACGACCGCGGAGCACCGGTGGCGTGGC 1054  
Qy  
576 -----ValAlaAlaSerAlaAspGl 582  
Db  
1053 CCCGGCCACCCCGGTGGTCAACCGGAGCAACCGGAGCGTATCGCGGTACCGCGGAACG 994  
Qy  
582 ySerArgPheValTrpAlaProGlyAspProGlyGlnProVal-----ValTyrAlaVa 600  
Db  
993 CGGCGCACCGGTGATCTCTACCGACGCGGCGGCAACCGATCGCGGACGCGTCAACCGCGCA 934  
Qy  
600 lGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValPro---AlaAsnAlaGlnIleAr 619  
Db  
933 CGGACGCGGCAAC---TGGAGCTTCAACCGCGCGCACGCGCTGGCCAAACGCGTCTCGGTAT 877  
Qy  
619 gSer-----As 621  
Db  
876 CAATGCGTGGCCAGGACGCGCGGCAACACCGAGCGCGCGCGCACCGACCGCGTGA 817  
Qy  
621 pArgValAsnProLysThr---PheTyrAlaLeuSerAsnGlyThrPheTyrArgSer--- 639  
Db  
816 CTCGGTAGCCCGCGCACCGCGGTGCTCGATCCGAGCAACGCGTACGGTATCAGCGGTAC 757

```
Qy 640 -----ThrAspGlyGlyValThrPheGlnProValAl 650
Db 756 CGCGGAAGCGGGGCCACCGGTGATCTCACCAGCGCGGC-----GGCAACCGGATACG 703
Qy 650 aAlaGlyLeu-ProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysG 670
Db 702 GCAGGGCACCGCGGATGGCAGCGGCACTGGAGCTTCACCTCCGGGCACA-CCGCTGACCA 644
Qy 670 luGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyHisSerThrAsnGly-----G 688
Db 643 ACGGCAGGTGATCATATGCGGTGGCCAGGACGCGCGGCAACACACGAGCGGTCCGGTCA 584
Qy 688 lySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysS 708
Db 583 GCACCAACAGTGGACGCGGTGGCCCGGCCACCCCGGTGATCGACCCGAGCAATGGTGTC 524
Qy 708 erAlaProGlySerSerTyProAlaValPheValValGlyThr-----IleGlyGlyV 726
Db 523 AACTCAGCGGCACCGCGGAACCGGGTCCGGGTGATCTCACCAGTGGCAATGGCAACC 464
Qy 726 alThrGlyAlaTyArgSerAspCysGlyThrTrpValLeuIleAsnAspAspG 746
Db 463 CGATCGGCCAGACCTCGCCGACGGTAGCGGC---AACTGGACCTTCACACCGGGCAGC 407
Qy 746 lnHisGlnTyArgAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgV 766
Db 406 CGCTGGCCAAACGCGCACGGTGGTCAACGCGTGGGCCAGGACCCGCGCGGC----- 357
Qy 766 alTyIleGlyThrAsnGlyArgGlyIleValTyGlyAspIleGlyGlyAla---ProS 785
Db 356 -----AATACAGCGGTCC-GGCCAGCACCCAGTGGATACGGTGGCGCGGCCACG 306
Qy 785 erGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerProS 805
Db 305 CCGGTGATCA-ATGCCA-----GCAACGGCAGCGGTGATCACCAGCACCG- 263
Qy 805 exProSerSerProSerProSerProSerProSerProSerProSerSerP 824
Db 262 --CCGAGGTGGCGCCAAAGTATCTCTCACCAGCGCAACGGCAACCCGATCGCGGAGAC 205
Qy 824 roSerPro-----SerProSerProSerProSerProSerProSerProSerProS 841
Db 204 CACCGCGCAGCGCAGTGGCAACTGGACCTTCACCCCGGCGCACGCCGCTGGCCCAACGGTAC 145
Qy 841 er---AlaSerProSerProSerProSerProSerProSerProSerProSerP 860
Db 144 GGTGATCAACGCCCTCGCCGAAG-----ACGCCGGGGCAACGCCAGCGGTCCGGC 94
Qy 860 roSerPro 862
Db 93 CAGCACCA 86
```

Search completed: May 14, 2004, 17:05:37  
Job time : 8674 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 14, 2004, 14:54:31 ; Search time 941 Seconds  
(without alignments)  
4615.210 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5135

Sequence: 1 MORSSEIRLTMRSRLVSL.....RASFGSVNPATPTADTYLOX 957

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2947324 seqs, 2269024515 residues

Total number of hits satisfying chosen parameters: 5894648

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09917376/runat\_14052004\_085258\_4727/app.query.fasta\_1.1095  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blotsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09917376@cgn\_1\_621@runat\_14052004\_085258\_4727  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09C\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
19: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
------------	-------	-------	--------	----	----	-------------

ALIGNMENTS

RESULT 1

US-09-917-376-2

; Sequence 2, Application US/09917376

; Publication No. US20040038334A1

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS

; FILE REFERENCE: 40197.4US01

; CURRENT APPLICATION NUMBER: US/09/917,376

; CURRENT FILING DATE: 2001-07-28

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 2869

; TYPE: DNA

; ORGANISM: Acidothermus cellulolyticus

; FEATURE:

; NAME/KEY: modified\_base

; LOCATION: (2869)

1	5134	100.0	2869	13	US-09-917-376-2	Sequence 2, Appli	
2	5134	100.0	2869	15	US-10-155-400-2	Sequence 2, Appli	
3	2487	48.4	9025608	15	US-10-156-761-1	Sequence 1, Appli	
4	2478	48.3	2646	15	US-10-156-761-1845	Sequence 1845, Ap	
5	1631.5	31.8	2517	15	US-10-026-994-4	Sequence 4, Appli	
6	1631.5	31.8	2710	15	US-10-026-994-1	Sequence 1, Appli	
c	7	1570	30.6	9025608	15	US-10-156-761-1	Sequence 1, Appli
8	1442	22.7	2217	15	US-10-156-761-2561	Sequence 2561, Ap	
9	1167.5	22.7	3668	10	US-09-927-827-21	Sequence 21, Appl	
10	1125.5	21.9	2646	13	US-10-395-241-11	Sequence 11, Appl	
11	1124	21.9	2481	13	US-10-395-241-17	Sequence 17, Appl	
12	1117	21.8	2367	13	US-10-395-241-13	Sequence 13, Appl	
13	940	18.3	5698	13	US-10-420-191-1	Sequence 1, Appli	
14	823	16.0	3687	10	US-09-917-384-2	Sequence 2, Appli	
15	823	16.0	3687	10	US-09-917-383-2	Sequence 2, Appli	
16	702	13.7	2289	10	US-09-917-378-2	Sequence 2, Appli	
17	361	7.0	2223	15	US-10-156-761-550	Sequence 550, App	
c	18	345	6.7	18596	9	US-09-880-107-1590	Sequence 1590, Ap
c	19	345	6.7	18596	9	US-09-967-768A-119	Sequence 119, App
c	20	345	6.7	18596	9	US-09-954-531-124	Sequence 124, App
c	21	345	6.7	18596	9	US-09-954-531-348	Sequence 348, App
22	330.5	6.4	7407	13	US-10-282-122A-30151	Sequence 30151, A	
23	330.5	6.4	7407	15	US-10-246-330-3	Sequence 3, Appli	
24	299.5	5.8	210528	13	US-10-087-192-289	Sequence 289, App	
25	297	5.8	32591	16	US-10-085-117-187	Sequence 187, App	
26	296.5	5.8	4767	10	US-09-940-316B-28	Sequence 28, Appl	
27	296	5.8	96602	16	US-10-085-117-61	Sequence 61, Appl	
28	293	5.7	4818	10	US-09-940-316B-32	Sequence 32, Appl	
c	29	292.5	5.7	185548	15	US-10-175-523-62	Sequence 62, Appl
c	30	289	5.6	324	13	US-10-282-122A-13296	Sequence 13296, A
31	288.5	5.6	4571	10	US-09-940-316B-18	Sequence 18, Appl	
32	288	5.6	288	16	US-10-437-708-111	Sequence 111, App	
33	288	5.6	288	16	US-10-395-402-111	Sequence 111, App	
34	285.5	5.6	5184	15	US-10-260-715-7	Sequence 7, Appli	
35	282.5	5.5	9903	13	US-10-282-122A-28185	Sequence 28185, A	
36	282	5.5	85692	17	US-10-461-194-1	Sequence 1, Appli	
37	281.5	5.5	432	9	US-09-864-761-32943	Sequence 32943, A	
38	281.5	5.5	518	9	US-09-864-761-16460	Sequence 16460, A	
39	280	5.5	86941	17	US-10-461-194-2	Sequence 2, Appli	
40	279	5.4	4547	10	US-09-940-316B-22	Sequence 22, Appl	
41	275	5.4	666	15	US-10-029-386-25948	Sequence 25948, A	
c	42	275	5.4	38918	15	US-10-017-161-2049	Sequence 2049, Ap
c	43	275	5.4	38918	16	US-10-292-798-1695	Sequence 1695, Ap
c	44	274.5	5.3	1168	15	US-10-017-161-2179	Sequence 2179, Ap
c	45	274.5	5.3	1168	16	US-10-292-798-1825	Sequence 1825, Ap

OTHER INFORMATION: a, c, t, g, other or unknown  
US-09-917-376-2

## Alignment Scores:

Pred. No.: 0 Length: 2869  
Score: 5134.00 Matches: 956  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.98% Indels: 0  
DB: 13 Gaps: 0

US-09-917-376-1 (1-957) x US-09-917-376-2 (1-2869)

```
Qy 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgLeuValSerLeuLeu 20
Db 1 ATGGATCGTTCGGAGAACATCGCTCTGACTATGAGATCACGACGATTGGTATCACTGCTC 60
Qy 21 AlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThr 40
Db 61 GCGGCCACTGCGTCTGTCGGCGGCGCGCTCTGGGAGTTCTGCCCATCGCGATAACG 120
Qy 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGly 60
Db 121 GCTTCTCTCCGCGCAGCGGGGAGACTACGCCGTACACTGGAGCAACGTGGCGATCGG 180
Qy 61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80
Db 181 GCGCGCGGCTTCTGCAAGCGGATCGTCTTCAATGAAGGTGCACCGGGAATCTGTAGCTG 240
Qy 81 ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu 100
Db 241 CGGACGACATCGGGGGGATGTATCGATGGATGCCGCCCAACGCGGGGTGGATCCCTCTT 300
Qy 101 LeuAsnTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp 120
Db 301 CTGGATTGGGTGGATGGAAATTTGGGGGTACAACGGCGTCTGACGATTGGCGGAGAC 360
Qy 121 ProIleAsnThrAsnIleValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPro 140
Db 361 CCGATCAATACTAACCAAGTATGGCGCGCGCTCGGAATGTACACCAACAGCTGGACCCA 420
Qy 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu 160
Db 421 AACGACGGAGCATTTCTCGCTCGTCTGATCAGGGCGCAACGTGGCAATAAAGCGCCCTG 480
Qy 161 ProPheIleLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180
Db 481 CCGTTCAAGCTTGGCGGCAACATGCCCGGCGGTGGAAATGGCGGAGCGGCTTGGGTGGAT 540
Qy 181 ProAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyIleValGlyLeuTrpArgSer 200
Db 541 CCAAAACAATGACAACATCTGTATTTTCGGCGCGCCCGCGGCGGCAAGGGCTCTGGAGAGC 600
Qy 201 ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle 220
Db 601 ACAGATTCCGGCGCGACCTGCTCCAGATGACGAACCTTCCGGAGCTTAGCACGATACAT 660
Qy 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla 240
Db 661 GCAAAATCCCACTGACAGACCGGCTATCAGAGCGGATATCAAGCGCTGCTGGGTGCT 720
Qy 241 PheAspIleSerSerSerLeuGlyGlnAlaSerIleThrIlePheValGlyValAla 260
Db 721 TTGCAAGTCTTCGTATCGCTCGGCGCAAGCAGATGAAGCAATTTTGGGGCGTGGCG 780
Qy 261 AspProAsnAsnProValPheTrpSerArgAspGlyAlaThrTrpGlnAlaValPro 280
Db 781 GATCCCAATTAATCCGGTCTCTGGAGCAGAGACGGCGCGGAGCTGGCAGGCGGTCCG 840
Qy 281 GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu 300
Db 841 GGTGCGCGGACCGGCTTCATCCGCGCAAGGGCGTCTTTGACCCGGTCAACACCGCTGCTC 900
```

```
Qy 301 TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpIlys 320
Db 901 TATATTGCCACCAAGCAATACGGGTGGTCCGTATGACGGGAGCTCCGGCAGCGTCTGGAAA 960
Qy 321 PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla 340
Db 961 TTCTCGGTGACCTCCGGGACATGACCGGAATCAGCCCGGTACCTTTCGACGACACGGCC 1020
Qy 341 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360
Db 1021 AACGACTACTTTGGTTACAGCGGCTCACTATCGACGCCAGCACCCGCAACACGATAATG 1080
Qy 361 ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly 380
Db 1081 GTGCAACCCAGATATCGTGGTGGCGGACACCAATATCTTTTCGGAGCACCGACGCGGT 1140
Qy 381 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 400
Db 1141 GCGACGTGGAGCGCGGATCTGGGATGGACGAGTTATCCCAATCGAAGCTTTCGATATGG 1200
Qy 401 LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro 420
Db 1201 CTTGACATTTCCGGGAGCGCTTGGCTGACCTTCGGGTACAGCCGAATCCTCCCGTACC 1260
Qy 421 SerProIysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 440
Db 1261 AGTCCGAAGCTCGGCTGGATGGAAGCGATGGCAATCGATCCGTTCAACTCTCATCGG 1320
Qy 441 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrIleThrAspSer 460
Db 1321 ATGCTCTACGGAACAGCGCGGACGTTGTACCAACAAATGATCTCAGGAAGTGGGACTCC 1380
Qy 461 GlyGlyGlnIleHisIleAlaProMetValIysGlyLeuGluThrAlaValAsnAsp 480
Db 1381 GCGCGCCAGATCATATTCGCGCCCGATGGTCAAGGATGGAGGAGACGCGCGTAAACGAT 1440
Qy 481 LeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 500
Db 1441 CTCATCAGCGCGCGCTTGGCGCGCGCTCATCAGCGCTCTCGGAGACCTCGCGCGCTTC 1500
Qy 501 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 520
Db 1501 ACCACCGCGAGCTTACTTGGCGTGCATCGACGATCTTCACTGTCACCGGTGTTCAGACC 1560
Qy 521 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 540
Db 1561 GGACACAGCTGCACTATCGGAATTTGAATCCGTCGATCATCGTTCGGCTGGAGTTTC 1620
Qy 541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyIysAsn 560
Db 1621 GATCCATCGAGCCAACCGAAGCAGACGACGCGCTCGGTTCTCGACAGACGCGCGCAAGAAC 1680
Qy 561 TrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla 580
Db 1681 TGGTTCCAAGCAGCAGAACTTGGCGGGGTGACGCGGCGGCGACCGTCCGCCATCGGCC 1740
Qy 581 AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTrpAlaVal 600
Db 1741 GACGGCTCTCGTTTCGTCTGGGCTCCCGCGCATCCCGGTACGCTGTGTGTACGACATC 1800
Qy 601 GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer 620
Db 1801 GGATTTGGCAACTCTCGGCTGCTTCGCAAGGTGTTCCCGCAATGCCAGATCCGCTCA 1860
Qy 621 AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr 640
Db 1861 GACCGGTGAATCAAAGACTTCTATGCCCTATCCAATGGAACTTCTATCGAAGCAGC 1920
Qy 641 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 660
Db 1921 GACGCGCGGTGACATTTCAAACCGCTCGCGCGCGGTCTTCGAGAGACGCGGTCCGCTCG 1980
Qy 661 ValMetPheHisAlaValProGlyLysGlyAspLeuTrpLeuAlaAlaSerSerGly 680
```



Db 1981 GTTCATGTTCCACGGCGCTTGAAGAAAGAGCGCATCTGGCTGCTGATCCAGCGGG 2040  
Qy LeuTyrHisSerThrAsnGlySerSerTyrAlaIleThrGlyValSerSerAla 700  
Db 2041 CTTTACCACTCAACCAATGGCGGAGAGTTGGTCTGCATCATCCGGGTATCTCCGGG 2100  
Qy ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 720  
Db 2101 GTGAACGTGGGATTTGGTAAGTCTGCGCGGGTGGTATACACCGCGCTTTGTGCTC 2160  
Qy GlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTyrVal 740  
Db 2161 GGACGATCGGAGCGGTACGGGGCGGTACCGCTCCGACGACTGTGGGACGACCTGGGTA 2220  
Qy LeuIleAsnAspAspGlnHisGlnTyrGlyAsnTyrGlyGlnAlaIleThrGlyAspHis 760  
Db 2221 CTGATCANTGATGACACGACCAATACGGAATTTGGGACAGCAATACCGGGTGACAC 2280  
Qy AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780  
Db 2281 GCGAATTTACGGCGGTGTATACATAGGACAGACGCGCGTGAATTTGTATACGGGACATT 2340  
Qy GlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeu 800  
Db 2341 GGTGGTGGCGCTCCGATCCGCTCCGTGGTGGTCCGTGGCTTCCGCGAGCGT 2400  
Qy SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820  
Db 2401 AGCCGAGCCGAGCCGAGCAGCTCCCATCCGCTCCGCTCCGCGAGCTCGAGTCCA 2460  
Qy SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840  
Db 2461 TCCTGTGGCGCTCTCGTCCGCTTACCATCGCCGAGTCCGTCTCGGTCTCCGTACCA 2520  
Qy SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860  
Db 2521 TCGCGCTCGCGAGCGCTCTTCGTACCCGAGCCCGCTTCGTCCACCGCTTCGTCCGCG 2580  
Qy SerProThrProSerSerSerProValSerGlyValLysValGlnTyrLysAsnAsn 880  
Db 2581 AGCCCAACGCGCTCGTCCGCGGTGTCGGGTGGGTGAGGTGAGTATAGAAATAAT 2640  
Qy AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900  
Db 2641 GATTTCGGCGCGGTGATATACATAGATCAAGCCGGGTTTGACGGGTGGTAATACCGGGTGG 2700  
Qy SerSerValAspLeuSerThrValThrValArgTyrTyrPheThrArgAspGlyGlySer 920  
Db 2701 TCGTGGTGGATTTGTTCGACGGTGACGGTGGGTGCTGCTGTTTCCCGGGATGGTGGCTCG 2760  
Qy SerThrLeuValTyrAsnCysAspTyrAlaAlaIleGlyCysGlyAsnIleArgAlaSer 940  
Db 2761 TCGACACTGGTGACAACTGTGACTGGCGGCGCATCGGGTGTGGGAATATCCGCGCTCG 2820  
Qy PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 956  
Db 2821 TTGGCTGGTGAACCGCGGACGCGCGACGCGGACACCTTACCTGCGAG 2868

## RESULT 2

US-10-155-400-2  
; Sequence 2, Application US/10155400  
; Publication No. US2003010898A1  
; GENERAL INFORMATION:  
; APPLICANT: ADNEY, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22

; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2869  
; TYPE: DNA  
; ORGANISM: Acidothermus cellulosoliticus  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (2869)  
; OTHER INFORMATION: a, c, t, g, other or unknown  
US-10-155-400-2

## Alignment Scores:

Pred. No.: 0 Length: 2869  
Score: 5134.00 Matches: 956  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.98% Indels: 0  
DB: 15 Gaps: 0

US-09-917-376-1 (1-957) x US-10-155-400-2 (1-2869)

Qy 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgArgLeuValSerIleLeu 20  
Db 1 ATGGATCGTTCCGAGAACATCCGCTGACTATGAGATCAGCAGCATTTGGTATCAGCTC 60  
Qy 21 AlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThr 40  
Db 61 GCCGCCACTGGCTGCTTCCGCGTGGCGCGCTCTGGGAGTTCTGCCCATTCGGATAACG 120  
Qy 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTyrSerAsnValAlaIleGly 60  
Db 121 GCTTCTCTCGCAGCGCGCGGAGACTCAGCGGTACACTGGAGCAACCTGGCGATCGGG 180  
Qy 61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80  
Db 181 GCGCGCGCTTTGTTCGACGGGATCGTCTTCAATGAAGGTGCACCGGGAATTTCTGTACGTG 240  
Qy 81 ArgThrAspIleGlyGlyMetTyrArgTyrAspAlaAlaAsnGlyArgTyrIleProLeu 100  
Db 241 CGGACGGACATCGGGGGGATGTATCGATGGGATGCCGCCAAGCGGGCGGTGATCCCTCTT 300  
Qy 101 LeuAspTyrValGlyTyrAsnAsnTyrGlyTyrAsnGlyValValSerIleAlaAlaAsp 120  
Db 301 CTGATTTGGTGGATGGAAACAATTGGGGGTACACACGGCTGTCAGCATTTGGCGGAGAC 360  
Qy 121 ProIleAsnThrAsnLysValTyrAlaAlaValGlyMetTyrThrAsnSerTyrAspPro 140  
Db 361 CCGATCAATACTAACAGGTATGGCGCGCTGCGGAATGTACACCAACAGCTGGGACCCA 420  
Qy 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTyrGlnIleThrProLeu 160  
Db 421 AACGACGAGCGATTTCTCCGCTCGTCTGATCAGGCGCAACGTGGCAATAACGCCCTCG 480  
Qy 161 ProPheIleLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180  
Db 481 CCGTTCAAGCTTGGCGGCAACATGCCCGGGCTGGAAATGGCGGAGCGGCTTCCGGTGGAT 540  
Qy 181 ProAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSer 200  
Db 541 CCNAAATGACACATTTCTGTTTTCGGCGCCCGGAGCGGCAAGGGCTCTGGAGAAGC 600  
Qy 201 ThrAspSerGlyAlaThrTyrSerGlnMetThrAsnPheProAspValGlyThrTyrIle 220  
Db 601 ACAGATTCGGCGGACCTGGTCCAGATGACGAATTTTCGGAGCTAGGACGATACATT 660  
Qy 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAla 240  
Db 661 GCAATCCCACTGACACGACCGGCTATCAGAGCGATATTCAAGGGCGTCTGTGGGTGGCT 720  
Qy 241 PheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAla 260  
Db 721 TTCGACAGTCTTCTGTCATCGCTCGGCGACGAGGAGTAGACCATTTTGTGGCGGTGGCG 780

261 AspProAsnAsnProValPheThrSerArgAspGlyGlyAlaThrTrpGlnAlaValPro 280  
781 GATCCCAATAAATCCGGTCTTCTGAGCAGAGACGGCGCGCGAGCTGGCAGGCGGTGCGC 840  
281 GlyAlaProThrGlyPheThrProHisLysGlyValPheAspProValAsnHisValLeu 300  
841 GGTGCGCCGACCGGCTTCATCCCGCAAGAGGGCGTCTTTGACCGGTCAACACGCTGCTC 900  
301 TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLys 320  
901 TATATTGCCACCAAGCAATACGGGTGTTCCGTATGACGGGAGCTCCGGCGACGCTCGAAA 960  
321 PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla 340  
961 TTCTCGGTGACCTCCGGGACATGGACGCGAATCAGCCCGGTACCTTCGACGGACACGGCC 1020  
341 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360  
1021 AACGACTACTTTGGTTTACAGCGGCTCACTATCGACCGCGAGCACCCGAAACACGATAATG 1080  
361 ValAlaThrGlnIleSerTrpTrpAspThrIleIlePheArgSerThrAspGlyGly 380  
1081 GTGGCAACCCAGATATCTGGTGGCCGCGACACCAATAATCTTTGGAGACCGACGGCGGT 1140  
381 AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal 400  
1141 GCGACGTGGACGGGATCTGGGATTCGACGAGTTATCCCAATCGAAGCTGGCGATATGTG 1200  
401 LeuAspIleSerAlaGluProTrpTrpLeuThrPheGlyValGlnProAsnProValPro 420  
1201 CTTTGACATTTCCGGCGGAGCCTTGCTGACCTTCGGCGGTACAGCGGAATCTCCCGTACCC 1260  
421 SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 440  
1261 AGTCCGAAGCTCGCTGGATGGATGAAGCGATGCAATCGATTCGTTCAACTCTGATCGG 1320  
441 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 460  
1321 ATGCTCTACGGACACAGCGCGGCGGTGTACGCAACAATGATCTCAGCAAGTGGGACTCC 1380  
461 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAsp 480  
1381 GCGCGCCAGATTATATCGCGCGCGATGGTCAAGGATTCGAGGAGACGGCGGTAAACGAT 1440  
481 LeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 500  
1441 CTATCAGCCCGCGCTTGGCGCGCCGCTCATCAGCGCTCTCGGAGACCTCGCGCGGCTTC 1500  
501 ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr 520  
1501 ACCCAGCGGACGTTACTGCGTGCCATCGACGATCTTCAGTCAACCGGTGTTCCAGCACC 1560  
521 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 540  
1561 GGCACACGCTCGACTATGCGGAATGAATCCGTCGATCATCGTTCCGCTGGAAGTTTC 1620  
541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 560  
1621 GATCCATCGAGCCCAACCGCAACGACGACGCTCGCGTTCTCGACAGACGGCGGCAAGAC 1680  
561 TrpPheGlnGlySerGluProGlyGlyValThrThrGlyThrValAlaAlaSerAla 580  
1681 TGGTTTCAAGCAGCGAACCTTGGCGGGGTGACGACGGCGGACCGTCCGCGCATTCGGCC 1740  
581 AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVal 600  
1741 GACGGCTCTGTTTCTGCTGGCTCCCGCGATCCCGGTGAGCTGTGGTGTACGCGATC 1800  
601 GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer 620  
1801 GGATTTGGCAACTCTCGGTGCTTTCGCAAGGTGTTCCCGCAATGCCAGATCCCGCTCA 1860

621 AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr 640  
1861 GACCGGTGATTCAAAGACTTTCTATCCCTATCAATGGAACCTTCTATCGAAGCAGC 1920  
641 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 660  
1921 GACGGCGGTGACATTCCAACCGGTTCGGCGCGGTCTTCGAGCAGCGGTGCGTCTCGT 1980  
661 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGly 680  
1981 GTCATGTTCCACCGGTGCTTGGAAAGAGCGCATCTGTGGCTCGTCATCGACGCGG 2040  
681 LeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAla 700  
2041 CTTTACCACCTCAACCAATGGCGGAGCATTTGGTCTGCAATCACCGCGGTATTCCTCCG 2100  
701 ValAsnValGlyPheGlyLysSerAlaProGlyLysSerSerTyrProAlaValPheValVal 720  
2101 GTGAACGTGGATTTGGTAACTCTGGCCCGGTCTCATACCGACCGCTCTTTGTCTGCTC 2160  
721 GlyThrIleGlyGlyValThrGlyValAlaTyrArgSerAspAspCysGlyThrThrTrpVal 740  
2161 GGCACGATCGGAGCGGTACGGGGCGGTACCGCTCCGACGACTGTGGGACGACCTGGGTA 2220  
741 LeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHis 760  
2221 CTGATCAATGATGACCGACCAATACGAAATTTGGGACAGCAATCACCGGTGACCAAC 2280  
761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780  
2281 GCGAATTTACGGCGGTGTATAGGACAGAACGGCGGTGGAAATGTATACGGGACAT 2340  
781 GlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeu 800  
2341 GGTGGTGGCGCTCGGATCGCGCTCTCGTGGGTGAGTCCGTCCGCTTCGCGGAGCTG 2400  
801 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820  
2401 AGCCGAGCGCGAGCGCGAGCAGCTCGCATCGCGTCCGCTCGCGAGCTCGAGTCCA 2460  
821 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840  
2461 TCCTCGTGGCGCTTCGCTCGCGTCCGCTCACCATCGCGAGTCCGCTCTCGGTCTCCG 2520  
841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860  
2521 TCGCGGTTCGCGACCGCGCTCTTCGTCACGAGCCGCTCTTCGTCACCGCTCTTCGTC 2580  
861 SerProThrProSerSerSerProValSerGlyValLysValLysValGlnTyrLysAsn 880  
2581 AGCCCAACCGCGTCTGCTCGTCCGCTCGGTGGGTGAGGTGAGGTGAGTATAGAATAAT 2640  
881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900  
2641 GATTCGCGCGCGGTGATTAATCAGATCAAGCGGGTTTGCAGTGGTGAATACCGGGT 2700  
901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 920  
2701 TCGTGGTGGATTTGTGACCGTGGCGGTGGTACTGGTTCCACCGGAGATGGTGGCTCG 2760  
921 SerThrLeuValTyrAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSer 940  
2761 TCGACACTGGTGTAACTGTGACTGGCGCGCATCGGGTGGGAATATATCCGCGCTCG 2820  
941 PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 956  
2821 TTCGGCTCGGTGAACCGCGGACCGGACCGCGGACACCTACCTCGAC 2868

## RESULT 3

US-10-156-761-1

; Sequence 1, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

```

; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
; US-10-156-761-1

Alignment Scores:
Pred. No.: 2,76e-166 Length: 9025608
Score: 2487.00 Matches: 490
Percent Similarity: 64.77% Conservative: 135
Best Local Similarity: 50.78% Mismatches: 248
Query Match: 48.43% Indels: 93
DB: 15 Gaps: 18

US-09-917-376-1 (1-957) x US-10-156-761-1 (1-9025608)

Qy 9 LeuThrMetArgSerArgLeuValSerLeuAlaAlaThrAlaSerPheAlaVal 28
Db 2277388 CTCACCGTGGAGAACCCGCACTTC-----ACGGCGGTGCTCGCGCTG 2277432

Qy 29 AlaAlaAlaLeu-----GlyValLeuProIleAlaIleThrAlaSerProAlaHisAla 46
Db 2277433 GCGCGCGGTCTGCCCGCGGACCCCGCGGACTGGCCGCGAGCGGCCCCACGCGGACG 2277492

Qy 47 AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyPheValAsp 66
Db 2277493 ATGCCCGCGGACACGTACAGCTGGAGAACCGCCGCGTCCGACGCGCGGCTTCGTCCCC 2277552

Qy 67 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 86
Db 2277553 GGCATCGTCTTCAACCGCTCCGAGAGAACCTCGCTACGCCCGCACCGACATCGCGCGC 2277612

Qy 87 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 106
Db 2277613 GCCTACCGCTGGCGCGGAGTCTCGAAGACCTGCGACGCGGCTGCTCGACTCGGTGCGCTG 2277672

Qy 107 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 126
Db 2277673 AGCGATGGGGGCACACGGGTGTGCGAGCTCGCTCGGCTCGGCTCGGCTCGGCTCGGCTCG 2277732

Qy 127 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146
Db 2277733 GTGTACCGCGCGCTCGGACGTACAGACGTGGACCGCGGACCGGCGGACCGGTCGCTC 2277792

Qy 147 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 166
Db 2277793 AGGTTCGGGCGACCGGGCGCGGAGCTGGCAGAACCGACCTGCGCTTCAAGCTGGGCGGG 2277852

Qy 167 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAspAsnIle 186
Db 2277853 AACATCGGGCGGGGCGATGGGCGAGCGGCTCGCGTTCGACCCGGAACAGGAACAGCGTG 2277912

Qy 187 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerGlyAlaThr 206

```

Db 2279878 CCG-----CATGTGGCTTCTCGACGGACACACGGCGCCCAACTGGTTCCGGGGGGCG 2279028  
Qy 566 GluProGlyValThrThrGlyGlyValAlaAlaSerAlaAspGlySerArgPhe 585  
Db 2279029 GACCTTCGGGGGTAGCGGGGTGGACGGTCCGGCGCGTCCGACGGCACTGCTTC 2279088  
Qy 586 ValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAenSer 605  
Db 2279089 GTGTGGAGCCGGCGGCGCACCGGG-----GTGCAGTACACGACCGGGTTCGGCACCTCG 2279142  
Qy 606 TrpAlaAlaSerGlnGlyValProAlaAlaAsnAlaGlnIleArgSerAspArgValAsnPro 625  
Db 2279143 TGTCTCGCGCTCCGGCGGCTCCCGCGCGGCGGATCTCGTCCGACCGGGTCGACCCG 2279202  
Qy 626 LysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThr 645  
Db 2279203 AGACCTTCTACGGCTTCAAGTCCGCGAGTGTCTACGTCACTCGGACGGGGCGGAC 2279262  
Qy 646 PheGlnProValAlaAla---GlyLeuProSerSerGlyAlaValGlyValMetPheHis 664  
Db 2279263 TTCACGGCGTCCGGCGCACCGGCTCCGAGCGCGCACAGC-----GTGCGCTTCAAG 2279316  
Qy 665 AlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSer----- 679  
Db 2279317 GCGCTCCCGCGACGAGGGCGGACATCTGGCTGGCGGGCGCGGACGCGCGGTAC 2279376  
Qy 680 GlyLeuTyrHisSerThrAsnGlyLysSerTrpSerAlaIleThrGlyValSerSer 699  
Db 2279377 GGGCTGTGGCACTCGACGAGCGGGCGGCGGCTTACCAAGTCCGCCACCGTCCAGCAG 2279436  
Qy 700 AlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 719  
Db 2279437 GCGGACACCATCGCTTCCGCAAGGGCGGCGACCGGCGCTCGTACGACGCTCTACACC 2279496  
Qy 720 ValGlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrTrp 739  
Db 2279497 AGCGCAAGATCGCGGTGTGCGCGGATCTTCGGTTCGACCGCACAGGGCGGAGCTGG 2279556  
Qy 740 ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyAlaIleThrGlyAsp 759  
Db 2279557 ACCCGGTCAACGACGATGCCACGATGGGTGGAGCGGCGGCGGATCACCGGTGAC 2279616  
Qy 760 HisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 779  
Db 2279617 CCAGGGTCTACGGCGCGGTATGTCTCACGACCGGGCGGATCGTACGGCGAC 2279676  
Qy 780 IleGlyAlaProSerGly-----SerProSerProSerValSerProSerAla 796  
Db 2279677 ACCCGGGTCTTTCGACGCGGCGGTACGAGCGCGGCGGACCGCGGCGCTGCACGGTG 2279736  
Qy 797 SerProSerLeuSer--ProSerProSerProSerProSerProSerProSerProSe 815  
Db 2279737 ACATACAGGATACGAAACAGTGGTGGCGGCGGCTTCCAGGCGGATGACAGTCCGCCAAC 2279796  
Qy 815 rProSerSerProSerProSerSer-----Pr 824  
Db 2279797 ACCGGGTGACCGCTGGGAGCGGTGTCTGGGTGGTCTGGTTCGGTACGGGCGAG 2279856  
Qy 824 oSerProSer----- 827  
Db 2279857 GTCAACAGCTGTGGAAAGCCTCGTACGCGGCGCGGTTTCGGGGGTGACCGGACGGAAC 2279916  
Qy 828 -----ProSerPro----- 830  
Db 2279917 CTGGCTTGGAAACGGGAGGTGGCGGCGGCTCTCGGTGAGCTTTCGGGTTTCACGGGAAGC 2279976  
Qy 831 -----SerProSerProSerProSerProSerProSerAla-----Se 843  
Db 2279977 TGTTCGGGATCCAAATGGCAGACCGACCGCTTCAAATCGCGGTGAGAACTGCACCGTTC 2280036  
Qy 843 rPro-----SerProSerSerProSerProSerProSerProSerProSerProSe 861  
Db 2280037 GCCTGATGCGAAGCGCA-CGCTCGCTGATCCGAGGGCGGCGGCGGCGGCGGCGGCGGCGG 2280095

Qy 861 rProThrProSerSerProValSerGlyGlyValLysValGlnTyrLysAsnAsnAs 881  
Db 2280096 CCCTCCGCTCTCTCACGGCGGTAGACCGGTGGCTG-----GGCAA 2280137  
Qy 881 pSer-AlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySerS 901  
Db 2280138 CTCCCGTGTCTGGTGGCAGGCTTCGGGAGTCCGGGCTCCGGTGGTGGCGGCGAGTA 2280197  
Qy 901 erSerValAsp 904  
Db 2280198 CCGGCGACGAC 2280208  
RESULT 4  
US-10-156-761-1845  
; Sequence 1845, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1845  
; LENGTH: 2646  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2646)  
US-10-156-761-1845

Alignment Scores:  
Pred. No.: 2,06e-169 Length: 2646  
Score: 2478.00 Matches: 473  
Percent Similarity: 64.07% Conservative: 128  
Best Local Similarity: 50.43% Mismatches: 227  
Query Match: 48.26% Indels: 110  
DB: 15 Gaps: 14

US-09-917-376-1 (1-957) x US-10-156-761-1845 (1-2646)

Qy 23 ThrAlaSerPheAlaValAlaAlaLeu-----GlyValLeuProIleAlaIleThr 40  
Db 22 ACGGCGGTCTCGCGGTGGCGGCGGTCTGCGCGCGGCGCACCCCGCGCACTGGCGCGG 81  
Qy 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGly 60  
Db 82 AGCG 141  
Qy 61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80  
Db 142 GCGCGCGCTTTCG 201  
Qy 81 ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu 100  
Db 202 CGCACCGACATCGGGCGCGCTACCGCTGGCGCGGAGTCTCTCAACCGCTCGGAGAACCTCGCCTAGCC 261  
Qy 101 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAsp 120  
Db 262 CTCGACTCGGTGGTGGAGCGGACTGGGGGCGACAGGGGTGTCTGTAGAGCTCGCTCCGAC 321



```
Db 2307 ----- 2307
Qy 854 SerSerProSerSerProSerProThrProSerSerSerProValSerGlyVal 873
Db 2308 -----GGTACGGAGCGCGCACCGACG-----GGCGCGTCG 2337
Qy 874 LysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnLysProGlyLeu 893
Db 2338 ACGGTGCATACAGGATCAGAACACAGTGTGCGGGCGCTTCCAG-----GCCGATGTA 2391
Qy 894 GlnValValAsnThrGlySerSerSerValAspLeuSerThrValThrValArgTyrTyr 913
Db 2392 CAGCTGCCCAACACCGGGGTGCACCGCTGGGACGGTGTGGTGGGC-----TGG 2442
Qy 914 PheThrArgAspGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaLysGly 933
Db 2443 TCGTTCGGTGACGGGACGAGGTGACCCAGCTGTGGNACGCTCGTACGGCGGCGCGGT 2502
Qy 934 CysGly-----AsnIleArgAlaSerPheGlySer 943
Db 2503 TCGGGGGGTGACCGCAGCAACCTGGCTGGGACCGGAGGGGTGGCGCGCGGTTC 2556

RESULT 5
US-10-026-994-4
; Sequence 4, Application US/10026994
; Publication No. US20030113732A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: EGVI Endoglucanase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC698
; CURRENT APPLICATION NUMBER: US/10/026,994
; CURRENT FILING DATE: 2002-04-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2517
; TYPE: DNA
; ORGANISM: Trichoderma reesei
US-10-026-994-4

Alignment Scores:
Pred. No.: 2,13e-108 Length: 2517
Score: 1631.50 Matches: 353
Percent Similarity: 57.23% Conservative: 130
Best Local Similarity: 41.82% Mismatches: 297
Query Match: 31.77% Indels: 64
DB: 15 Gaps: 23

US-09-917-376-1 (1-957) x US-10-026-994-4 (1-2517)
Qy 25 SerPheAlaValAlaAlaLeuGlyValLeuProfileAlaIleThrAlaSerProAla 44
Db 10 TCTCGAGTCCTGCCCTGTGCTGGGGCGGTC-----ATCCCTGCC 51
Qy 45 HisAlaAlaThrThrGlnProTyrThrThrThrThrSerAsnValAlaIle-----GlyGlyGlyGly 63
Db 52 CATGCTGCC-----TTTCATGGAAGAACGTCAGCTCGGGCGGCGCGGCG 99
Qy 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp 83
Db 100 TTCGTCCCCGGGCATCATCTTCCATCCCAAGACAAAGCGGTAGCATATGCAACGACAGAT 159
Qy 84 IleGlyGlyMetTyrArgTyrAspAlaAlaAsnGlyArgTyrIleProLeuLeuAspTyr 103
Db 160 ATTGGCGGGCTGTACCCCTCAAC-----GCCGACGACTCATGGACCGCGGTACCGGATGGG 216
Qy 104 Val-----GlyTyrAsnAsnThrGlyTyrAsnGlyValValSerIleAlaAla 119
Db 217 ATTGCTGATAATGCGGCTGGCACAACTGG-----GGCATCGAGCGCTGTGGCGCTT 267
```

```
Qy 120 AspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp 139
Db 268 GATCCGCGAGGACGATCAAAAGGTGTATCGCGCAGTCGTCATGATATACGAAACAGCTGGGAT 327
Qy 140 ProAsnAspGlyAlaIleLeuAtqSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 159
Db 328 CCAGTAAATGGAGCCATCATTTCTGCTCGTCAGACCGCGGCGCAAGCTGTCTCTTCAACCAAC 387
Qy 160 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 179
Db 388 TTGCGCTTCAAGTCCGGGGTAAATGCCAGGACCGGAGCGGAGAGCGTCTGGCTGTC 447
Qy 180 AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArg 199
Db 448 GATCGCGCCAACTCCAAACATCATCTTTGGTGTCTGCTCAGGAACAGCGCTCTGTGAAG 507
Qy 200 SerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyr 219
Db 508 TCTACGGACCGCGCGGTGACCTTTTCAAGGCTCTCGTGTTCACGGCAACTGGGACGTAC 567
Qy 220 IleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpVal 239
Db 568 ATCCAGACCGAGTGATTCACCGGTACACAGCGCAACAGCAGAGGACTCATGTGGGTT 627
Qy 240 AlaPheAspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyVal 259
Db 628 ACGTTGACTCAACACGACGACGACCGGGGAGCCAGCTCTCGTATCTTTGTGGCAGC 687
Qy 260 AlaAspPro---AsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAla 278
Db 688 GCTGATAACATCACTGCTTCACTGTGTGAGACGAATGCCGCTCCACCTGGAGTGTCT 747
Qy 279 ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis 298
Db 748 GTACCGGGGCGAGCGCGGAAATCTTCTCACAGGCGCAAACTGCACCGCAGCAGAGAG 807
Qy 299 ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspVal 318
Db 808 GCCTTGTATCTGACCTATTCCGATGGCACAGGCGCGTATGATGGCACACTTGGCTCAGTG 867
Qy 319 TrpLysPheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAsp 338
Db 868 TGGAGGTACGACATTCGAGGGGGAACCTGGAAAGACATCAACCCCTGTCTCTGGATCAGAT 927
Qy 339 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr 358
Db 928 CTA-----TACTTTGGCTTTGGCGGCGCTTGGCCCTCGATTTGCAAAAGCCAGGAACC 978
Qy 359 IleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAsp 378
Db 979 CTGTGTGTGTCTTGTGAACCTCTTGGTGGCCAGATGCTCAGCTGTTCGTTCGACCGAC 1038
Qy 379 GlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArg 398
Db 1039 TCTGGGCAACATGAGGCCCGATCTGGCGGTGGCGAGCTATCCGACTGAGACCTATTATAC 1098
Qy 399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGly---ValGlnProAsnPro 417
Db 1099 TACAGCATCTCAACTCCCAAGCACCGGTGATCAAGAACAACTTTATCGATGTGACGAGC 1158
Qy 418 ProValProSer-----ProLysLeuGlyTrpMetAspGluAlaMetAlaIle 433
Db 1159 GAGTCACGTCGAGTGTCTCATCAGGCGCTCGGCTGGATGATGATGCTCTCCAGATT 1218
Qy 434 AspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsn 453
Db 1219 GACCAACCGACAGCAACCACTGGCTCTACGGCAGCGGAATGACAACTCTTTGGCGGCCAC 1278
Qy 454 AspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeu 473
Db 1279 GATCTCCCAACTGGGACACGCGCAACATGTGTCAATCCATCACTGCGACGCGCATC 1338
```



QY 474 GluGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAla 493  
Db 1339 GAGAAATTCCTCGTCAGGACCTGGGCTCTGCAACCGCGGAGAGCTATTATGGCGCA 1398  
QY 494 LeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValProSer 510  
Db 1399 GTCGAGAGACACACCGCTTCACCTTGGCAGGAAACACCTCGGGACATCGCCGAG 1458  
QY 511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 530  
Db 1459 ACGTCTGGGCAACGCCACATGGCCACCTCGACGAGCTCGACTACGCCGGGAATCG 1518  
QY 531 ProSerIleValAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 550  
Db 1519 GTCAAGAGCGTCTCGCGCTCGCAACCGCGGACGCAACAG----- 1563  
QY 551 ValAlaPheSerThrAspGlyLysAsnTyrPheGlnIleSerGluProGlyGlyVal 570  
Db 1564 GTGGCCATCTCGTCCAGCGCGCGAGCTGGAGCATCGACTACCGCGCGACAGTCC 1623  
QY 571 ThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTyrAlaProGly 590  
Db 1624 ATGAAGCGCGGACGGTGGCTTATCGCGCAGCGGACGACGATCTCTGTCGACCGCC 1683  
QY 591 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTyrAlaAsnSerGln 610  
Db 1684 TCGTCCGCG-----GTGCAGCGCTCGCAGTTCAGGGGAGCTTTCCTCGTCTCG 1734  
QY 611 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProIysThrPheTyrAla 630  
Db 1735 AGCTCTCGCGCGCGCGCTCATCGCTCGCAAGAAAGACCAACACGCTCTTCTACGCC 1794  
QY 631 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla 650  
Db 1795 GGGTCTCGGATCGACCTTTTACGTGCAAGAGCACCGCGGAGCGCTTC-----ACG 1845  
QY 651 AlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValProGly 668  
Db 1846 CGCGGG---CCCAAGTGGGCGGCGGAGGACGATCGGGATATCGTGTCTACCCGACC 1902  
QY 669 LysGluGlyAspLeuTyrPheAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGly 688  
Db 1903 ACCGCGGCGACGTTGATGTCTCGACGCGCTCGGATATTCGCTCCACAGACTCGGGC 1962  
QY 689 SerSerTyrSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707  
Db 1963 ACCACTTTGGCCAGTCTCCACCGCCCTGACCAACACTTACAGATCGCCCTGGGTGTG 2022  
QY 708 SerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyValThr 727  
Db 2023 GGTCTCA---GGCTCGAATGG---AACCTGTATGCTTCCGACCC-----GGCCCGTCA 2070  
QY 728 GlyAla-----TyrArgSerAspAspCysGlyThrThrTyrValLeuIleAsnAspAsp 745  
Db 2071 GGGGCTCGCTCTACGCCAGTGAGAGACAGCGCGCTCTCGGAGGACATCCAGGGCTCC 2130  
QY 746 GluHisGlnTyrGlyAsnTyrGlyGlnAlaIleThrGlyAspHisAlaValLeuArg 765  
Db 2131 CAGGGCTTCGGCTCCATCGACAGCACCAAGGTGCGCGGCGGACGCGCGCGGCGCA 2190  
QY 766 ValTyrIleGlyThrAsnGlyArgGlyIleValTyr-----GlyAspIleGlyGlyAla 783  
Db 2191 GTCTAGTGGGACCAACCGCGCGGGCGTCTTTTACGCTCAGGGAACCGTCCGCGCGCGC 2250  
QY 784 ProSerGlyProSerProSerProSerValSerProSerAlaSerProSerLeuSerProSer 803  
Db 2251 ACGGCGGGACT-----TCTCTCGTCGACCAAGCAGAGC 2283  
QY 804 ProSerProSerSerSerProSerProSerProSerProSerProSerSerSerSer 823  
Db 2284 AGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCCAGCTGAGGTGAGCGTGTATCCACG 2343  
QY 824 ProSerProSerProSerProSerProSerProSerProSerProSerProSerAlaSe 843

Db 2344 ACCCGGGCTTCGACGAGTGTGACTTCGTGCGAGGACCACTCGGCGCGGTCCTCCACGGGGTCA 2403  
QY 843 rProSerPro 846  
Db 2404 GGGGTGCGCG 2413  
RESULT 6  
US-10-026-994-1  
; Sequence 1, Application US/10026994  
; Publication No. US20030113732A1  
; GENERAL INFORMATION:  
; APPLICANT: Dunn-Coleman, Nigel  
; APPLICANT: Goedegebuur, Frits  
; APPLICANT: Ward, Michael  
; APPLICANT: Yao, Jian  
; TITLE OF INVENTION: ECVI Endoglucanase and Nucleic Acids  
; TITLE OF INVENTION: Encoding the Same  
; FILE REFERENCE: GC698  
; CURRENT APPLICATION NUMBER: US/10/026,994  
; CURRENT FILING DATE: 2002-04-30  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2710  
; TYPE: DNA  
; ORGANISM: Trichoderma reesei  
US-10-026-994-1  
Alignment Scores:  
Pred. No.: 2,31e-108 Length: 2710  
Score: 1631.50 Matches: 353  
Percent Similarity: 57.23% Conservative: 130  
Best Local Similarity: 41.82% Mismatches: 297  
Query Match: 31.77% Indels: 64  
DB: 15 Gaps: 23  
US-09-917-376-1 (1-957) x US-10-026-994-1 (1-2710)  
QY 25 SerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44  
Db 50 TCTCGAGTCTTGTGCTTGTCTGCGGGCCGCT-----ATCCCTGCC 91  
QY 45 HisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGly 63  
Db 92 CATCTGCC-----TTTTCATGGAAGAACGTCACAGCTCGGCGCGCGCGCGC 139  
QY 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp 83  
Db 140 TTCGTCCCGGCATCATCTTCATCCCAAGACAAAGGGGTAGCATATGCACGACAGAT 199  
QY 84 IleGlyGlyMetTyrArgTyrAspAlaAlaAsnGlyArgTyrIleProLeuLeuAspTyr 103  
Db 200 ATTGGCGGGGTGTACCGGCTCAAC---GCCGACGACTCATGACCGCGCTCGCGATGG 256  
QY 104 Val-----GlyTyrAsnAsnTyrGlyTyrAsnGlyValValSerIleAlaAla 119  
Db 257 ATTCTGATTAATGCGGCTGGCAACATGG-----GGCATCGACCTGTGGCGTT 307  
QY 120 AspProIleAsnThrAsnLysValTyrAlaAlaValGlyMetTyrThrAsnSerTyrAsp 139  
Db 308 GATCCGAGGACGATCAAAAGTGTATGCGCGAGTCGCGCATGTATACGACAGCTGGAT 367  
QY 140 ProAsnAspGlyAlaIleLeuAsnSerSerAspGlnGlyAlaThrTyrGlnIleThrPro 159  
Db 368 CCGAGTAATGAGGCATCATCTCGTCTGTCAGACCGCGCGCAACGCTGCTTCCAAAC 427  
QY 160 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 179  
Db 428 TTGCCCTTCAAGTCGGGGGTAAATCCACGAGCGCGGAGCGGAGCGTCTGGCTGTC 487  
QY 180 AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArg 199



488 GATCGGCGCACTCCAAACATCATCTTCTGCTCGCTCAGGAAACGGCCTCTGGAAG 547  
200 SerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyr 219  
548 TCTACGAGCGGCGGTGACCTTTTCAAGGTCTCGTTCACGGCAACTGGGACGTAC 607  
220 IleAlaAsnProThrAspThrGlyTyrGlnSerAspIleGlnGlyValValTrpVal 239  
608 ATCCAGACCCGAGTATCCCAACGGCTACAACAGCAAGCAAGCAAGCAAGTCTATGTTGGT 667  
240 AlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyVal 259  
668 AGCTTCGACTCAACAGCAGCAGCAGCGGGGAGCCACCTCTGTAATCTTTGTTGGCAGC 727  
260 AlaAspPro---AsnAsnProValPheThrSerArgAspGlyAlaThrTrpGlnAla 278  
728 GCTGATAACATCACTGCTTTCAGTCTATGTGAGCAGCAAGATGCGGCTCCAGTGGAGTCT 787  
279 ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis 298  
788 GTACCGGGGCGCAGCGGGAATCTTCTCTCAACAGCGCAACTGCGAGCCAGCAGAGAAG 847  
299 ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspVal 318  
848 GCCTTGTAATCACTGATTTCCGATGTCAGCGGCGCTGATGATGGCAGCTTGGCTCAGTG 907  
319 TrpLysPheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAsp 338  
908 TGGAGGTACGACATTCAGGGGGAATCTTGGAAAGACATCAACCTCTCTCTGGATCAGAT 967  
339 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr 358  
968 CTA-----TACTTTGGCTTTGGCGGCTTGGCTCGATTTGCAAAAGCCAGCAAGACC 1018  
359 IleMetValAlaThrGlnIleSerTrpThrAspThrIleIlePheArgSerThrAsp 378  
1019 CTTGTGTGTGCTTTTGAACCTTGGTGCCAGATGCTCAGCTGTTTGGTGCAGCCGAC 1078  
379 GlyGlyAlaThrTrpThrArgIleTrpAspThrTrpSerTyrProAsnArgSerLeuArg 398  
1079 TCTGGGACACATGAGCCGATCTGGGCGTGGCGAGCTATCCGACTGAGACTATTAC 1138  
399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGly---ValGlnProAsnPro 417  
1139 TACAGCATCTCAACTCCCAAGCAGCGTGGATCAAGAACAACTTATCGATGTGACGAGC 1198  
418 ProValProSer-----ProLysLeuGlyTrpMetAspGluAlaMetAlaIle 433  
1199 GAGTCACCGTCCGATGGTCTCATCAAGCGCCTCGGCTGGATGATGATGCTCTCGAGATT 1258  
434 AspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsn 453  
1259 GACCAACCGCAGCAGCAGCAGCTGCTCTACGACCGGAGTGAATCTTTGGCGGCCAC 1318  
454 AspLeuThrLysTrpAspSerGlyGlnIleHisIleAlaProMetValLysGlyLeu 473  
1319 GATCTCACCACATGGGACACGCGCCCAATGTGTCAATCAATCACTGGCAGCAGCGCATC 1378  
474 GluGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAla 493  
1379 GAGGAATTCCTCGCTCAGGACCTGGCTCTGCACCGCGGGAAGCAGGCTATTGGCGCGCA 1438  
494 LeuGlyAspLeuGlyPheThrHisAla-----AspValThrAlaValProSer 510  
1439 GTCCGAGACGACACCGCTTCACTTTGGCCAGCAGAAACGACCTCGGAGCATCGCCGCGAC 1498  
511 ThrIleThrSerProValPheThrGlyThrSerValAspTyrAlaGluLeuAsn 530  
1499 AGCTGTGGGCAACGCCACATGGGCCACTTCAGCAGGCTCGACTACGCGGGAAGTCTCG 1558  
531 ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 550  
1559 GTCAAGAGCGTCTCTCGCGTGGCAACACACCGCGGCGCAGCAAG----- 1603

QY 551 ValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyVal 570  
Db 1604 GTGCCATCTCGTCCGACGGCGCGCATCGATCGATCGACGCGCCGACGACGTC 1663  
QY 571 ThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGly 590  
Db 1664 ATGACCGCGCGCACGGTGGCTTATCGGCCGACGACGATCTCTGTGTCGACGCGC 1723  
QY 591 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 610  
Db 1724 TCGTCCGCG-----GTGACGCGCTCGAGTTCAGGGCAGCTTTGGCTCCGCTCG 1774  
QY 611 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla 630  
Db 1775 AGCTGCGCGCGCGCGCTCATCGCTCGCAAGAACCAACAGAGCTCTTCTACGCGC 1834  
QY 631 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla 650  
Db 1835 GGTCTCCGATCGACCTTTTACGTCAAGCAAGCACCGCGCAGCAGCTTC-----ACG 1885  
QY 651 AlaGlyLeuProSerSerGlyAlaValGlyValMet-----PheHisAlaValProGly 668  
Db 1886 CGCGGG---CCAGCTGGGCGCGCAGCAGTCCGGGATATCGCTGCTCACCGGACC 1942  
QY 669 LysGluGlyAspLeuTrpLeuAlaAspSerGlyLeuTyrHisSerThrAsnGlyGly 688  
Db 1943 ACCGCGGCGACGTGTATGTCTGACCGACGCTCGGCATATTCGCTCCACAGACTCGGCG 2002  
QY 689 SerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707  
Db 2003 ACGACCTTTGGCAAGTCTCCACCGCTCCCAACACTACAGATCGCTCGGTGGTGTG 2062  
QY 708 SerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyValThr 727  
Db 2063 GGCTCA---GGCTGGAAGTGS---AACTGTATGCTTCGGCACC-----GGCCGCTCA 2110  
QY 728 GlyAla-----TyrArgSerAspCysGlyThrTrpValLeuIleAsnAspAsp 745  
Db 2111 GGGGCTCGCTCTACGCGTGGAGACAGCGCGGCTCTCTGGACGACATCCAGGGCTCC 2170  
QY 746 GlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg 765  
Db 2171 CAGGCTTCGGCTCCATCGACAGACCAAGGTCCGCGCAGCGCAGCAGCCCGGGCAA 2230  
QY 766 ValTyrIleGlyThrAsnGlyValGlyIleValTyr-----GlyAspIleGlyAla 783  
Db 2231 GTCTACGTGGCACCAACGGCGGGGCTCTTTTACGCTCAGGAAACGCTCGCGCGCGC 2290  
QY 784 ProSerGlySerProSerProSerProSerProSerProSerProSerProSerProSer 803  
Db 2291 ACGGGCGGACT-----TCCTCGTCGACCAAGCAGAGC 2323  
QY 804 ProSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 823  
Db 2324 AGCAGCAGTACTCTCTCGCCAGCTCGAGCAGCAGCTGAGGTGTGTATCCAGC 2383  
QY 824 ProSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 843  
Db 2384 ACCCGGCTTCGACGGTGTCTGTCGAGGACACGCTCGCGCGCGCTCCACGGGGTCA 2443  
QY 843 rProSerPro 846  
Db 2444 GGGGTCGCGC 2453

## RESULT 7

US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030115018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
PRIORITY FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIORITY FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIORITY FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 1  
LENGTH: 9025608  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (4187715)  
OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Alignment Scores:  
Pred. No.: 3,65e-100 Length: 9025608  
Score: 1570.00 Matches: 393  
Percent Similarity: 49.25% Conservative: 131  
Best Local Similarity: 36.94% Mismatches: 346  
Query Match: 30.57% Indels: 194  
DB: 15 Gaps: 38

US-09-917-376-1 (1-957) x US-10-156-761-1 (1-9025608)

Qy 4 SerGluAsnIleArgLeuThrMetArg-----SerArgArgLeuValSerLeu 19  
Db 3162026 TCTCCAGAACGAGGTATGTCATCGCAGCCGCCCGCGGACGACGACGACGTC----- 3161973  
Qy 20 LeuAlaIleThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIle 39  
Db 3161972 CTCGCGGGACCGCGCG-----GCCGCGCGTCTCAGCGCGTCCCG-----GCC 3161928  
Qy 40 ThrAlaSerProAlaHisAlaAlaThr-----GlnProTyrThrTrpSerAsn 56  
Db 3161927 GTGGGTCCAGGCGCAGCGCGCGGAAACACCGCGCGCGTCTGTCACCGTGGCGAAC 3161868  
Qy 57 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 76  
Db 3161867 GCCGTATCGGGGACCGCGTCTCAGCGCGTCTCTTCCACCGCTCAGCGCGT 3161808  
Qy 77 IleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArg 96  
Db 3161807 CTCGCTACCGCGGACCGACATCGCGCGCGCTTACCGTGGGACGACCGCGCGCGC 3161748  
Qy 97 TrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSer 116  
Db 3161747 TGGACCGGTCTCATGACCACTCGGTGGGACGACTGGAACCTCTCGGGTTCGAGCG 3161688  
Qy 117 IleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsn 136  
Db 3161687 ATGGCGGTTCGACCCAGCACCGCGCGCTCTTACCTCGCGTGGGACCGCTACCGCCAG 3161628  
Qy 137 SerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGln 156  
Db 3161627 TCGTGGCGGGGACAC-----GGCGGGTCTCTCGCTCCGAGACCGCGCGCGACCTGGACC 3161571  
Qy 157 IleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArg 176  
Db 3161570 CGCACCACTGACCGTGAAGTCTCGCGGCGCAACGAGACCGCGCGCGCGGTGAGCGA 3161511  
Qy 177 LeuAlaValAspProAsnAsnAsnAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 196  
Db 3161510 CTCCTCGTCGACCGCGCGCAGCGACACCTCTGGCTGGGACG-----CGCACGACGCG 3161454

Qy 197 LeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVal 216  
Db 3161453 CTGCTCAAGTCGACCGCGGGCGCCACCTTGGGACCGCGCGACCCCTTCCCG----- 3161400  
Qy 217 GlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVal 236  
Db 3161399 -----GGAGGGCGAATCTCTCCGGGCGAGGAGTC 3161370  
Qy 237 ValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePhe 256  
Db 3161369 GTGTTC-----CTCGTCGCGCGCGCGCGACCGTCTAC 3161337  
Qy 257 ValGlyValAlaAspProAsnAsnPro-----ValPheTrpSerArgAsp 271  
Db 3161336 GCCGCTCGGGTACCGGACCGCACCTCGGCGACCGCGAACTGTACCGCAGCGCGAC 3161277  
Qy 272 GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe-----IlePro 288  
Db 3161276 ---GGCAGACCTGGGGGCGCGTCCCGCGCGCGCTCCGGCACCTCCCGCAAGGTCTCCG 3161220  
Qy 289 HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 308  
Db 3161219 CTCGCGCGCGGTACGACGACGACCGCGCGAGCTGTAGTCAGTCAGCGCGACGCC 3161160  
Qy 309 GlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrp 328  
Db 3161159 GGCGCGGGCGGACAGTCGACGCGCGTGTGACAGTCGTCAGCGCGCGGACGTGG 3161100  
Qy 329 ThrArgIleSerProVal-----ProSerThrAspThrAlaAsnAspTyrPhe 344  
Db 3161099 ACCGAGGTCAACCGCGTGAAGCGCGCGGACGACGAGCGAGCGGTCTCGCGCGACCTTC 3161040  
Qy 345 GlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGln 364  
Db 3161039 GCCTACGCGGGGTGCGGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3160980  
Qy 365 IleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThr 384  
Db 3160979 AACCGTGGCGCGCGGACGCGTCTTCCGTCACGCGCGCGCGCGCGCGCGCGCGCG 3160920  
Qy 385 ArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSer 404  
Db 3160919 TCCCTCAAGGAC-----GCCGCGGTGTTCGACGCTGTC 3160887  
Qy 405 AlaGluProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeu 424  
Db 3160886 GAGACTCCCTTCTCGACTGGGCGCGAC-----AAGCGGAATTC 3160845  
Qy 425 GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGly 444  
Db 3160844 GCCTGGTGGTCCAGCGCTCGCGTGCACCGTACCGTCCAGCACGCTGCTGTACGGG 3160785  
Qy 445 ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIle 464  
Db 3160784 ACCGCGCGACCTCTACGCGACCGCGACCTCAAGCGGTG----- 3160743  
Qy 465 HisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerPro 484  
Db 3160742 -----GCACCGGATCCGCGCGCTGGAGGAGAGCGCGCTCGCGCAACTGATCTCGCCC 3160689  
Qy 485 ProSerGly---AlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAla 503  
Db 3160688 CCGTGGGGGAGGACACCTGATACGCGGACCTCGGGGACATCGGTGTGATGTACACGAG 3160629  
Qy 504 AspValThrAlaValProSerThr---IlePheThrSerProValPheThrThrGlyThr 522  
Db 3160628 CGGCTCACGGGTCTCCGTCGCGCGCATGGCGACGACCGCGCTGTCGGTCTCGCGACG 3160569  
Qy 523 SerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspPro 542  
Db 3160568 GGACTCGCGGACCGCGCGCGCGGTATGTCTCGCGACGCGCGCTGGGCGCGACCC 3160509  
Qy 543 SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp--- 561



Qy	27	AlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAla	46
Db	52	-----GCCGCGCGCTCACGGCGTCC-----GCCGTGGGGTCCAGCGCACGCC	99
Qy	47	AlaThrThr-----GlnProTyrThrTrpSerAsnValAlaIleGlyGlyGly	63
Db	100	GCGAAACCAACCGCGGCCCGTGTACGTGGCGCAACCGCGTCATCGGGGGGACCGGC	159
Qy	64	PheValAspGlyIleValPheAsnGluAlaProGlyIleLeuTyrValArgThrAsp	83
Db	160	TTGTGACCGCGGTCTTCCACCCCTCGGTACGGGTCTCGCTACGCCCGGACCGAC	219
Qy	84	IleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp	103
Db	220	ATCGCGCGCGCTACCGCTGGGACGCGCGCGCCCGCTGGACCCCGCTCATCGACCA	279
Qy	104	ValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsn	123
Db	280	CTCGCTGGGACGACCTCTCGGCGTCGAGCGATGGCGGTCTGACCCCGCACCGC	339
Qy	124	ThrAsnIleValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGly	143
Db	340	CGGACCGGCTCTACCTCGCGCTGGGACCTACCGCCAGTCTGGCGGGCAAC--GGC	396
Qy	144	AlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLys	163
Db	397	CGGTCTCGCTCCGAGGACGCGCGCCACCTGGACCCGACCGCTGACCGTGAAG	456
Qy	164	LeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgIleAlaValAspProAsnAsn	183
Db	457	CTCGCGCGCAACGAGGACGCGCGCGCGCTGAGCGACTCTCTCGTCGACCCCGCGAC	516
Qy	184	AspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSer	203
Db	517	AGCAGACACCTCTCGCTGGGCAAG--CGGCACGACGGGTCTCAAGTCGACCGACGG	573
Qy	204	GlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnPro	223
Db	574	GCGCCACTGGGACGCGCGACCGCTTCCG-----	606
Qy	224	ThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLys	243
Db	607	-----GCGAAGGCGAACTCTCCGCGGCGGAGTCGTGTTTC-----	642
Qy	244	SerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAlaAspProAsn	263
Db	643	-----CTCGTCCGCGCGCGCGCACCGCTCTACGCGCGCTGGGGTGACCGCGAC	690
Qy	264	AsnPro-----ValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAla	278
Db	691	GGACCTCGGGCACGGGAACCTGTACGACGCGCGAC--GGCAGACCTGGGGGGCC	747
Qy	279	ValProGlyAlaProThrGlyPhe-----IleProHisLysGlyValPheAspPro	295
Db	748	GTCCCCGCGCGCCCTCCGGCACCTCCGCAAGTCCCGTCCGCGCGCGGTACGACAG	807
Qy	296	ValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSer	315
Db	808	CACACCCCGAGCTGTACGTACGACGCGGACGACCCCGCGCGCGCGCCAGTCCGAC	867
Qy	316	GlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIleSerProVal---	334
Db	868	GCGAGCTGCAACAGTCGTACCGCCACGGGGACGTGGACCGAGGTACCCCGGTGAAG	927
Qy	335	-----ProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIle	351
Db	928	CCGGCGGGACGACGACGCGGCTCGCGCGGACACCTTCGCTACGCGCGGGTCTCGCGTC	987
Qy	352	AspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThr	371
Db	988	GACCGCGCGCGCGCACCTCTGCTCTCCACCAACACCGCTGGCGCGGACCGCGAC	1047

Qy	372	IlellePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSer	391
Db	1048	ACGGTCTTCGGGTCCACGACGGCGCGGTACCTGCACGTCCTCCACGAGC	1098
Qy	392	TyrProAsnArgSerLeuAArgTyValLeuAspIleSerAlaGluProTrpIleuThrPhe	411
Db	1099	-----GCCGCGTGTTCGAGCGTTCGAGACTCCCTTCCTCGACTCG	1140
Qy	412	GlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGluAlaMet	431
Db	1141	GGCGACGAC-----AAGCCGAGTTCGGCTGGTGGATCCAGCGCGTC	1182
Qy	432	AlaIleAspProPheAsnSerAspArgMetLeuTyGlyThrGlyAlaThrLeuTyAla	451
Db	1183	CGCGTCACCGTACGACTCCACACACTCGTGTACGGACCGCGCGACCTCTACGGC	1242
Qy	452	ThrAsnAspLeuThrLysTrpAspSerGlyGlyIleHisIleAlaProMetValLys	471
Db	1243	ACCGCGACCTCAAGCGCTGG-----GCACCGCGATCCCGC	1278
Qy	472	GlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSerGly---AlaProLeu	490
Db	1279	GGCTCGAGAGAGCGCGCTGCGCCAACTGATCTCGCCCCGTGCGGGAGGCACACCTG	1338
Qy	491	IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer	510
Db	1339	ATCAGCGGACTCGGGACATCGGTGTATGTACCAGAGCGGCTCACGGCGTCTCCGTG	1398
Qy	511	Thr---IlePheThrSerProValPheThrGlyThrSerValAspTyThrAlaGluLeu	529
Db	1399	CGCGCATGCGAGCAACCCCGTGTTCGGTTCGGCGACCGGACTCGCGCAGGCGCGGCC	1458
Qy	530	AsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg	549
Db	1459	AGCGCGCGTATGTGTCGCCGACGGCTGGGGCGACACGGCAACGGC-----	1506
Qy	550	HisValAlaPheSerThrAspGlyGlyLysAsnTrp-----PheGlnGlySerGluPro	567
Db	1507	-----GCGTACTCCACGACGCGGGCGGACCTGGGGCGCTTCGAGGCCACAGCCGAC	1560
Qy	568	GlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrp	587
Db	1561	ATCGCCAAAGACGACCGCGGACCGATCGCCACAGTGCAGCGCGCGGCACACTGCTGTG	1620
Qy	588	Ala-----ProGlyAspProGlyGlnProValValTyAla-----	599
Db	1621	TCCTTCGTCACGTGGGACGGC-----ACGACGTACGGCGCCACCGCTCG	1665
Qy	600	ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg	619
Db	1666	ACGACAAACGGCGGAGTGTTCGAGGTCTCCTCTCCCGAAGGGCGCCACACCGGTC	1725
Qy	620	SerAspArgValAlaAsnProLysThrPheTyAla-----LeuSerAsnGlyThrPheTy	637
Db	1726	GCGGACCGCGCGCATCCGACGCGCTTCTACGCTACGACTTCGACATGGAACGCTATAC	1785
Qy	638	ArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGly	657
Db	1786	GCACGACTACAGTGGCGCTGTTTCACGGCCCGTGGCGGGGAGCTGCC-----TCGGC	1842
Qy	658	AlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAla	677
Db	1843	GACAGCCAGTTCAAGCTGGTTCGGCGCGCGGACGAGCGGACCTGTGGTCTCCGCC	1902
Qy	678	Ser---SerGlyLeuTyHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGly	696
Db	1903	AAATGGAACGGGTCTACTCCGCTCCACGACGGCGGGGACACCTTCGCCAGGATCGACAG	1962
Qy	697	ValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyTrpProAla	716
Db	1963	TGTCGGCGCTCGTACCTCGCTTCGGCAAGCGCGCGGACGCGCGGACCTACCGCGGC	2022
Qy	717	ValPheValValGlyThrIleGlyGlyValThrGlyValArgSerAspAspCysGly	736

```

Db      2023  ATCTACAGGTGGCTCGACGAGACCATCACCGCGCTCTACCGCTCCGACGACGCGCC 2082
Qy      737  Thr-Thr-TripValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 756
Db      2083  AGACATGGTGGCTCGGATCAACGACGACGCCACACAGTGGGGTGGATCGCGGAGCGCTC 2142
Qy      757  ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 776
Db      2143  GTCGGTGACCGCGCATCCACGCGCGGTCTACTCTGCCCAACCAACGCGCGGCATCCAG 2202
Qy      777  TyrGlyAsp 779
Db      2203  TACGGGGAG 2211

RESULT 9
; Sequence 21, Application US/09927827
; Publication No. US2003003617eA1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramsauer, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)/B
; CURRENT APPLICATION NUMBER: US/09/927,827
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 21
; LENGTH: 3668
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(2668)
US-09-927-827-21

Alignment Scores:
Pred. No.: 9,15e-75 Length: 3668
Score: 1167.50 Matches: 299
Percent Similarity: 48.55% Conservative: 137
Best Local Similarity: 33.30% Mismatches: 351
Query Match: 22.74% Indels: 112
DB: 10 Gaps: 28

US-09-917-376-1 (1-957) x US-09-927-827-21 (1-3668)
Qy      23  ThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSer 42
Db      1136  ACCACAACAGGAGCATGGCTGGCTGTCTTGAGCCTGCTG---CTGTGCTCTTTTCGACG 1192
Qy      43  Pro-----AlaHisAlaAlaThrGlnProTyrThrTrpSerAsnValAla 58
Db      1193  CCCAGCGTGGCTGGCGCGGACGCCGACGCTCGGGGCTTACCAGTGGCGAGTGTGCGCC 1252
Qy      59  IleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeu 78
Db      1253  ATTGGCGGTGGCGCTTTGTTCACGGGTGTGCTGTGTTCATCCCGCGCAACGTGGTCTGGCC 1312
Qy      79  TyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIle 98
Db      1313  TATCGCGCACCGCATGTGGGTGGCGGTACCGCTGGATGGCAGCGCAGCAGTGGACC 1372
Qy      99  ProLeuLeuAspTrpValGlyTyrAsnAsnTrpGlyTyrAsnGlyValValSerIleAla 118
Db      1373  CGCGTACCGACCTGTTGGCGGTGACGACCTGGAACTGTATGGGCATCGACGATCGCC 1432
Qy      119  AlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrp 138
Db      1433  GTCGACCCCGCATGCGCATGGCTGTATCTGGCGCGCGCACCTATATGCAAGAA--- 1489
Qy      139  AspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThr 158
```

```

Db      1490  CGCGCCGGGCACTGCCGCGAGTGTGGCTCTCAACCGCGCGCGACGCTTCAGCGCTGCC 1549
Qy      159  ProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGlyArgLeuAla 178
Db      1550  GACCTGGCTTTAAGCTGGGTGTAAACAGCTGGGCGCGGCCAATGGCGAGCGGCTGGCG 1609
Qy      179  ValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrp 198
Db      1610  GTGACCGCGCATCGCGCGGTGTCTGTCTGGCTCGCGCATGCC---GGCTGTGG 1666
Qy      199  ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 218
Db      1667  CGTAGCGAGCATCGCGCGCGCACTGGCGGAAGTGGCGTCTTTCGAGACGCCGCGCTG 1726
Qy      219  TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp 238
Db      1727  GCCGTTGCCACCGCGCGCAATCATGTTGGGCGGAGCAGCGGTG---GGGATCGCTTT 1783
Qy      239  ValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGly 258
Db      1784  GTCGTGTTTCGACGCGCAGTGGCAACACGCTCGCCCAACACCGCGCATCTACGTGGCG 1843
Qy      259  ValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAla 278
Db      1844  GTGTCCACCGAACAACACGCGCTGTATGTCTCGAAGATGCCGCGCGCATTTGGGCACCG 1903
Qy      279  ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis 298
Db      1904  GTGCGCGGCAACACGCGCGCTGGCGCCGAGCCACATGGCGCGGCGGACGATGGGCAC 1963
Qy      299  ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerGlyAspVal 318
Db      1964  ---TGGTATCTGAGCTATGGCGACCGCGCGCGGCGGACCTGATGCGCGGGGAGCGCTTG 2020
Qy      319  TrpLysPheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAsp 338
Db      2021  TGGAAATTCAGCGCGCACAGCGCGCTGGCGTGAGATCAGCCGATTCGCGAG---CCA 2077
Qy      339  ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr 358
Db      2078  GCCAGTGGCGAGTGGATTCGGTGGGTGGGTGGGTGGATCCGCAACATCCGAGGTG 2137
Qy      359  IleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAsp 378
Db      2138  CTGCTGGCGACGACCTTCGCGCGTCGACGCGCGCGGACGAGCTGTATCGACGCTGGAT 2197
Qy      379  GlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArg 398
Db      2198  GGTGGCAAGCATGGACGCGCTG-----TTGGCGCGAT 2230
Qy      399  TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProPro 418
Db      2231  CGCGTGTTCGATCAACGCGCGCGCGGTGG---ACCGCATGCGCCGCGCGAC----- 2281
Qy      419  ValProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSer 438
Db      2282  -----TGGATGGGG---GGCTGGCGATCGATCGTTCGACGCG 2317
Qy      439  AspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrp 458
Db      2318  AACCATGCGCTGTTCTGACCGGCTACGCGCATCTGGCGCTCGCGCATCTG----- 2368
Qy      459  AspSerGlyGlyGlnIleHisIleAlaPro-----MetVal 470
Db      2369  -----CAGGATTTCCGCGCACCGCGCGCGCGCTCGAGTGGTGGTTCAGGAC 2416
Qy      471  LysGlyLeuGluThAlaValAsnAspLeuIleSerProProSerGlyAlaProLeu 490
Db      2417  CGTGGCTGGAGAAACCGTGGCGCTGGACCTGTCTCAGCCCGATGGCTGGCGGCGCATCTG 2476
Qy      491  IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 510
```

Db 2477 CTAGCGCGCTCGCGATATCGAGCGCTTCGGCATCAGCACTGGACCGGTG---CAG 2533  
Qy 511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 530  
Db 2534 TTGCAGTACCGCGCGCCACCGCTGACCAATGTGGCAAGCATCGATCGCGCGCCAGCGC 2593  
Qy 531 ProSerIleValAlaGlySerPhe-AppProSerSerGlnProAsnAspArgHi 550  
Db 2594 CCGCAGTGGGTGGCGAGCGGTACCGTGGCGCGCGCGCAACACAGAAATCCGCGCG 2653  
Qy 550 sValAlaPheSerThrAspGlyGlyLeuAsnTrpPheGln---GlySerGluProGlyGlu 569  
Db 2654 C---TGTAATACCGGTGATCGCGCAAGCAATGCTGCTTTCGCGAGTGGCGCGCAGC 2710  
Qy 569 yValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPr 589  
Db 2711 CGGCGAGCGCGCGCAGCATTCGCTTGTGGCATCGCGCGCAGTGTGTGGGACC 2770  
Qy 589 oGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSe 609  
Db 2771 GCGCGCGCGCGCAAT---TGGCGACGTCGCACTTCGCGCGCAGTGGCAGCGTGT 2824  
Qy 609 rGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTy 629  
Db 2825 GAACGCGCTGCCGAACCCCGCGTGGTGTGACCGATCGGTGGACACGCGCGCTGGTA 2884  
Qy 629 rAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGl 647  
Db 2885 TCGGTGGATGTCCGACGCGGCGAGCTGTACGAGACGACCGATCCCGCGCGAGTTTCG 2944  
Qy 647 nProValAlaAlaGlyLeu----- 653  
Db 2945 T-----GCGACCGGTGTACAGTGGCGCAGCGCGCGCATGAGCGCACCGCGCGCA 2998  
Qy 654 -----ProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGl 671  
Db 2999 GCTCGTCCGCGACCGCGCGCGCGCGTGTGTATCTCGCCAGCGCCCGCGCAAG----- 3053  
Qy 671 yAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTr 691  
Db 3054 -----GGTGTGATGCGCTGGCGAGACGCT-----GCGCT 3082  
Qy 691 pSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGl 711  
Db 3083 GCAGTGTCTACACCGCGCAAGCACGCTCGCTGGGATCGGCAAGCGCTCGCGGC 3142  
Qy 711 ySerSerTrpAlaValPheValGlyThrIleGlyGlyValThrGlyAlaTyrAr 731  
Db 3143 CCGTGGCGCGCGCGCTGTATCTGGCGCGCGCGTGCAGCGCGTGTGATGGCGGTTCG 3202  
Qy 731 qSerAspAspCysGlyThrTrpValLeuIleAsnAspGlnHisGlnTyrGlyAs 751  
Db 3203 CTCGACGATGGCGCGCGCGCGAGCGCATCAACGATGACGCGCACCGCTTCGCGGC 3262  
Qy 751 nTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAs 771  
Db 3263 C---CGTACAGCGTACCGGTGATCGCGCATTCGCGCGCGTGTACTTCGCCACCG 3319  
Qy 771 nGlyArgGlyIleValTyrGlyAspIleGlyAlaProSerGlySerProSerProSe 791  
Db 3320 CGGCGCGCGCATTTCTACGCGCATCCGAGATGAGCAGTGCATGCTGTCGCGA----- 3374  
Qy 791 rValSerProSerAlaSerProSerLeuSerProSerProSerProSerSerSerPro-- 810  
Db 3375 ---CGTCTGCTGCTTTACCGTACGCGCGCGCGCGCTGCGCGTGGCGCGCGCGT 3430  
Qy 811 -----SerProSerProSerProSerProSerProSerProSerProSer-- 827  
Db 3431 ACACCGCGACCTGGAGATCACCGATGACCCCGT---CCACATTTCCCGCGCTGCTG 3489  
Qy 828 -----ProSerProSerProSerProSerArg----- 836  
Db 3490 CTGGGTGTGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTTT 3549

Qy 837 -----SerProSerProSerAlaSerProSerProSerProSerSerSerPr 850  
Db 3550 GCCGATGGCGCGGTGCTGCAGCGGATGAGGCAATGCCGGTATGGGGCTGGGCCACGCC 3609  
Qy 850 oSerProSerSer-----SerProSerSerSerProSerProSerPro 864  
Db 3610 GCGCGCGCAATCAGCGTGAGTTTCGATGCAAGCAGCGCCACGCCAAGACCG 3661  
RESULT 10  
US-10-395-241-11  
; Sequence 11, Application US/10395241  
; Publication No. US20040038367A1  
; GENERAL INFORMATION:  
; APPLICANT: YAOI, Katsuro  
; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME  
; FILE REFERENCE: Q73756  
; CURRENT APPLICATION NUMBER: US/10/395,241  
; PRIOR APPLICATION NUMBER: JP 2002-83433  
; PRIOR FILING DATE: 2002-03-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 11  
; LENGTH: 2646  
; TYPE: DNA  
; ORGANISM: Geotrichum sp. M128  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (120)..(2558)  
US-10-395-241-11

Alignment Scores:  
Pred. No.: 6,89e-72 Length: 2646  
Score: 1125.50 Matches: 287  
Percent Similarity: 49.82% Conservative: 131  
Best Local Similarity: 34.21% Mismatches: 294  
Query Match: 21.92% Indels: 127  
DB: 13 Gaps: 32

US-09-917-376-1 (1-957) x US-10-395-241-11 (1-2646)

Qy 17 ValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuPro 36  
Db 114 TTAGATATGGTGGCGATCCCTCCCTCGGGAAGCGCTTACTGCGCTTTTCGATTCTG--- 170  
Qy 37 IleAlaIleThrAlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsn 56  
Db 171 -----GCTCGCTCGCGTGGCAAGGACACTACGAGTTCAAGAT 212  
Qy 57 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 76  
Db 213 GTCCGATCGCGCGCGGTACATTACCGGATGTGTCGCGACCCCAAGACCAAGGAC 272  
Qy 77 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArg 96  
Db 273 CTGCTGTACGCGCGCACGACATTTGGCGCGCGTACCGCTGGGACGCGACGTCGAAG 332  
Qy 97 TrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSer 116  
Db 333 TGGATCCCGCTCAACGACTTTATCGAGCGCGGACATGAACATTATGGGACCGAGTGC 392  
Qy 117 IleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThr--- 135  
Db 393 ATCGCGTGGACCCCAACACCCGACAGCGGTGTACTCGCGCAGGGCGCTATGTCGCG 452  
Qy 136 AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrp 155  
Db 453 GACGAGTGG-----GCGCGCTTCTATGTGTCGCAAGACCGCGCGCGCTGTTT 500  
Qy 156 GlnIleThrProLeuProPheLeuGlyGlyAsnMetProGlyArgGlyMetGlyGlu 175







;; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
 ;; FILE REFERENCE: 073756  
 ;; CURRENT APPLICATION NUMBER: US/10/395,241  
 ;; PRIOR FILING DATE: 2003-03-25  
 ;; PRIOR APPLICATION NUMBER: JP 2002-83433  
 ;; PRIOR FILING DATE: 2002-03-25  
 ;; NUMBER OF SEQ ID NOS: 19  
 ;; SOFTWARE: PatentIn version 3.2  
 ;; SEQ ID NO 17  
 ;; LENGTH: 2481  
 ;; TYPE: DNA  
 ;; ORGANISM: Artificial Sequence  
 ;; FEATURE:  
 ;; OTHER INFORMATION: Xyloglucan Oligosaccharide-Degrading Enzyme with Histidine Tag  
 ;; FEATURE:  
 ;; NAME/KEY: CDS  
 ;; LOCATION: (1)..(2481)  
 US-10-395-241-17

Alignment Scores:  
 Pred. No.: 8,25e-72 Length: 2481  
 Score: 1124.00 Matches: 298  
 Percent Similarity: 48.91% Conservative: 127  
 Best Local Similarity: 34.29% Mismatches: 306  
 Query Match: 21.89% Indels: 138  
 DB: 13 Gaps: 33

US-09-917-376-1 (1-957) x US-10-395-241-17 (1-2481)

```

Qy 52 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValaspGlyIleValPheAsn 71
Db 13 TAGGAGTTCAGAGATGTCGGATCGCGGCGGGGTACATTACCGGGATGTGCGGCAC 72

Qy 72 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 91
Db 73 CCAAGACCAAGACCTGCTGTACGGCGCGCACGACATTGGCGGCGGTACCGCTGGGAC 132

Qy 92 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 111
Db 133 GCAGGACGTCCTCAAGTGGATCCGCTCAACGACTTATCAGGCGCGAGGACATGAACATT 192

Qy 112 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 131
Db 193 ATGGGACACGATCGATCGCGTGGACCCCAACACCCGACGCGTGTACCTCGCGCAG 252

Qy 132 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 150
Db 253 GGGCGCTATGTCGGCGACGAGTGG-----GCGGCGTTCATGTGTCGAAGAC 300

Qy 151 GluGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 170
Db 301 CGCGGCGCAGTCGTTTCAATCTACGATGTCGCGTTCGCCGATGGCGGCAACGACATGGGA 360

Qy 171 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 190
Db 361 CGCAACATGGCAGCGCTCGCTGTCACCCCTTCACTCGAACGAGGTCGTGATGGGT 420

Qy 191 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 210
Db 421 ACGCGTACA---GAGGATATCTGGAAGAGTTCGAGCCGCGCCCAAGACCTTGACAAAGCTC 477

Qy 211 ThrAsnPheProaspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 230
Db 478 ACCTTCATCCGCGCGGTTTC-----ACCAACGGTATCGGATACACG 519

Qy 231 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGln 250
Db 520 TCG-----GTCAATTTTCGACCCC-----GAA 540

Qy 251 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 270
Db 541 CGTAATGGCACCATCTACGGGCGGCGACTGCCCCCGCAGGGC---AGTACGTACCGCAC 597
  
```

```

Qy 271 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle----- 287
Db 598 GACGCGGTGTCTCGTGGGAGCCAGTGGCGGCGCCAGCGCTCCAGCTGGCTCAACAGGACC 657

Qy 288 -----ProHisLysGlyVal 292
Db 658 ACGGCGCGTTCCCGGACAAAGCCCGCTGATCGCGCCGACGCCATGAAGTCGCT 717

Qy 293 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 312
Db 718 CTCACCCCC-----AATCTCTCTACGTGACTTACGCGGACTACCTGCTCATGCGGC 771

Qy 313 GlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle--- 331
Db 772 GTCAGTTCGGCAAGTCTGGCGCAGAACCGCACTCGGGCGCTCGGGACGACATTACT 831

Qy 332 -----SerProValProSerThrAspThrAlaAsnAspTyrPheGly 345
Db 832 CCCGCGTCGGCAACTCGCTCGCTGCCCTGACCAACACGACGCTTCCTCGCGCGGA 891

Qy 346 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 365
Db 892 TTTTGGCGTCTCAGCGTCGACGCGACCAACCCCAACCTCTCGTCTCATCACC---CTC 948

Qy 366 SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 382
Db 949 GACCGCGACCCGCGACCGCCCTCGACAGCATCTACCTCTCAACCGATGCCGCGCGGACC 1008

Qy 383 TrpThrArgIleTrpAspTrpTrpSer-----TyrPro 393
Db 1009 TGGAGGAGCGTCAACCGAGTCTCGTCCCGTCCAACTCGAAGGTAACTGGGGCCACCG 1068

Qy 394 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyVal 413
Db 1069 ACTAACCGCGCGCGGTAC---AAGGACGCGCACGCGCTGTTCGTGGCTCGACTTCAACAAC 1125

Qy 414 GlnPro-----AsnProProValProSerPro-----LysLeuGly 425
Db 1126 GGTCCCGATGGGGGGATACGTTGCGCGCGACGTCGCGCGCTCCCGGCTCACCAAGTTGGC 1185

Qy 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445
Db 1186 TGGTGGATGAGCGTGTCTTATCGATCCGTTCAACCCCGAGCACCTGATGATACGCGCAG 1245

Qy 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly 461
Db 1246 GGGCGGACCATCTGGCGGACCGCACGCTCTCCCGTGTGAGAGAGACTGG----- 1296

Qy 462 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluGluThrAla 477
Db 1297 -----GCGCCGAGTGTGATCTCCAGATCGACGCGTATCGAGAGATGCG 1341

Qy 478 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 497
Db 1342 ATCTGTGCTCCGCTCGCCCAAGCGCGCGCGCTCTCTGCGGACATCGGTGACATT 1401

Qy 498 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 517
Db 1402 AGCGGATGAAGCAGCAGCAGCTCACCAAG---CCCCAGAAAGATGTTTGGTGGCGCCAG 1458

Qy 518 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 537
Db 1459 TTCTCCAACTTCGACGATCGACGCTCGCGGCAACTTCCCGACGTTGTCTCGCGCGC 1518

Qy 538 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 557
Db 1519 GGATCTCGGGACACGAGTACGACGCGGTGGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 1578

Qy 558 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrThr----- 572
Db 1579 GGAGACGCGTGGACCATCTTCTCCCTACCTGCGCTCTCTGCGCATGAACGCGGACCATACCAG 1638
  
```



QY 288 -----ProHisLysGlyVal 292  
Db 655 ACGGGCGGTTCCGGCAAGAACCCGCGTGCATCGCGCGCAGCCCGCAAGTGCCT 714  
QY 293 PheAspProValAenHisValLeuTyrIleAlaThrSerAenThrGlyGlyProTyrAsp 312  
Db 715 CTCACCCCC-----AACTTCCTCTAGTGACTTACCGCCGACTACCTCGTGGTCCATGGGGC 768  
QY 313 GlySerSerGlyAspValTyrLysPheSerValThrSerGlyThrTyrThrArgile--- 331  
Db 769 GTCACGTTCCGGGAAGTCTGGCGCCAGAACCGCAGCTCGGGCGCTGGGACGACATTACT 828  
QY 332 -----SerProValProSerThrAspThrAlaAenAspTyrPheGly 345  
Db 829 CCCCGCTCGGCAACTCGTCGCTCGCCCTGACCAACACAGAGCTTCCCTCGCGGGGGA 888  
QY 346 TyrSerGlyLeuThrIleAspArgGlnHisProAenThrIleMetValAlaThrGlnIle 365  
Db 889 TTTTGGCGGTCTACGCGTGCAGCGGACCAACCCCAACCGTCTCGTGCATCACC---CTC 945  
QY 366 SerTyrProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 382  
Db 946 GACCGCACCCCGGACCGCCCTCGACAGCATCTACCTCTCAACCGATCGCGGCGGAC 1005  
QY 383 TrpThrArgIleTyrAspThrIleThrSer-----TyrPro 393  
Db 1006 TGGAGGACGTCACCCAGCTCTCGTCCCGTCCAACTCGAAGCTAACTGGGGCCACCG 1065  
QY 394 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTyrLeuThrPheGlyVal 413  
Db 1066 ACTAACCGCGCGGTAC---AAGGACGCGACCGCTCTTCCGTGGTCTGACTTCAACAAC 1122  
QY 414 GlnPro-----AsnProProValProSerPro-----LysLeuGly 425  
Db 1123 GGTCCCGAGTGGGGGGGATACGGTGGCGCGGACGCTACGCGCGGCTTCAACAACTTGGC 1182  
QY 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445  
Db 1183 TGGTGGATGAGCGTGTGCTTATCGATCCGTTCAACCCCGAGCACCTGTGATGTACGGCAG 1242  
QY 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThys-----TrpAspSerGly 461  
Db 1243 GGGGCGACCATCTGGGCGACCGACACGCTCTCCGCTGTCGAGAAGGACTGG----- 1293  
QY 462 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluThrAla 477  
Db 1294 -----GGCGCGAGCTGTACTCCAGATCCAGCTACCGGCTTCAACAACTTGGC 1338  
QY 478 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 497  
Db 1339 ATCTGTGCTCGCTCGCCCAAGAGCGCGCGGCGTCTCTGCGGCATCGGTGACATT 1398  
QY 498 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 517  
Db 1399 AGCGGATGAAGCAGCAGCAGCTCACCAAG---CCCCAGAAGATGTTTGGTGGCGCCCGAC 1455  
QY 518 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 537  
Db 1456 TTTCTCAACCTCGACAGCATCGACGCTGCGGGCAACTTCCCAACCGTTGTGCTCGCGGCC 1515  
QY 538 GlySerPheAspProSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 557  
Db 1516 GGATCTCGGGACAGATACGACGCGGTGGCGCGGCGGTGCTGACGCGTACGCGC 1575  
QY 558 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThr----- 572  
Db 1576 GGAGACGCTGGACCATCTTCCCTACTCTGCTCTGTCATGAACGCGGACCTACCAG 1635  
QY 573 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro-----Gly 590  
Db 1636 GGCAGCAGCATTTGACGTCGACGCGGCGGCGGCGGACGATGCTGTGTCGACCAAGCTTGAC 1695  
QY 591 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 610

Db 1696 GAGCAGGCTCGGGACCGTGGTACTCGCACGACTATGGCAAGACGTGGTCT----- 1746  
QY 611 GlyValProAla-----AenAlaGlnIleArgSerAspArgValAsn 624  
Db 1747 ---GTTCCCGCTGGCGACCTGAAGCCCGAGACTGCAATGTCTCTCGACCAAGGTCAG 1803  
QY 625 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVal 644  
Db 1804 GATGSCACGTTCTACGCTACCGATGGCGCAAGTCTTCTGCTCTCGACCGACGCGGGAAG 1863  
QY 645 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 664  
Db 1864 TCGTATGCGCGCAAGGGCGCGGACTTGTCACT-----GGCACATCGCTCATGCCCTGCC 1917  
QY 665 AlaValProGlyLysGluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHis 683  
Db 1918 GTGAACCCCTGGTGGCGCGGACGCTCTGGGTGCTGTTCGCGAGGGCGTCTTCTCCAC 1977  
QY 684 SerThrAenGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 702  
Db 1978 TCGACCGACTTGGCGCTCTGTTACAGAGGTAGGTACCGCCACCGACCTCTCGTGGC 2037  
QY 703 ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe 718  
Db 2038 GTCGCGCGCCCAAGTCCAAAGTCGACGCAAGAGGCTAGCGCGCTCCCGGCTTCTTC 2097  
QY 719 ValValGlyThr-----IleGlyValThrGlyAlaTyrArgSerAspCysGly 736  
Db 2098 ATCTGGGGCACCGCAAGCTGGAAAGCAGCATCGGCTGTACCGCTCCGACGACACGCG 2157  
QY 737 ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 756  
Db 2158 AGCAGCTGGAGCGGCTCAATGACGAGGACAACTACTCGGGC---CCCACCATGATC 2214  
QY 757 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 776  
Db 2215 GAGCGCGACCCCAAGGCTCTCGGGCGCTGTATCTAGGACGACGCGCGGTCGTCG 2274  
QY 777 TyrGlyAspIle 780  
Db 2275 TACGCGGACCTT 2286

RESULT 13  
US-10-420-191-1  
; Sequence 1, Application US/10420191  
; Publication No. US20040067569A1  
; GENERAL INFORMATION:  
; APPLICANT: No. US20040067569A1ozymes Biotech, Inc.  
; APPLICANT: Rey, Michael W.  
; APPLICANT: Zaretsky, Elizabeth J.  
; APPLICANT: Haas, Jeffrey A.  
; TITLE OF INVENTION: Polypeptides Having Xyloglucanase Activity And Nucleic Acids  
; FILE REFERENCE: 10210-200-US  
; CURRENT APPLICATION NUMBER: US/10/420,191  
; PRIOR FILING DATE: 2003-04-18  
; PRIOR APPLICATION NUMBER: US 60/373,987  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1  
; LENGTH: 5698  
; TYPE: DNA  
; ORGANISM: Tricoderma reesei  
US-10-420-191-1

Alignment Scores:  
Pred. No.: 3,71e-58 Length: 5698  
Score: 940.00 Matches: 374  
Percent Similarity: 30.11% Conservative: 136  
Best Local Similarity: 22.08% Mismatches: 315  
Query Match: 18.31% Indels: 871





Qy 711 GlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla--- 729  
Db 4413 AGTCCGAGCTTGACTTTGGACATACGATCGCAC---GGCCGCTCGAGGGCTCGC 4466  
Qy 730 ---TyrArgSerAspAspCysGlyThrThrTyrVal----- 740  
Db 4467 CTCTAGCCAGTGGAGACACGGCGCTCTCTGAGAGCCGTGGCCGGCAGTCCCGAGCG 4526  
Qy 741 -----LeuIleAsnAspAspGlnHisGlnTyrGly--- 750  
Db 4527 GAGATCGGTCACCTCTGTCTGCGCGCGGAGGACCTCGGACATCCAGGCTCCAGGGCTTC 4586  
Qy 750 ----- 750  
Db 4587 GGCTCCATCAGACACCAAGTTCGCGCGCAGCGGCTGTAGTTCGAGGGTCCCGAAG 4646  
Qy 751 -----AsnTyrGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 766  
Db 4647 CCGAGGTAGTGTCTGTGTTCCAGCGCGCTGCGCGCAGCACCGCCGG---CAAGTC 4700  
Qy 767 TyrIleGlyThrAsnGlyArgGlyValValTyrGlyAsp----- 779  
Db 4701 TAGTGGGACCAACGCGCGGGGCTCTTTACGCTCAGGCGTCTGCGCGCGCTTCAG 4760  
Qy 780 -----IleGlyGlyAlaPro 784  
Db 4761 ATGCACCGCTGGTTCGCGCGCCCGCAGCAAAATGCGAGTCCGAAACGTCGCGCGCGC--- 4817  
Qy 785 SerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerPro 804  
Db 4818 ACGGGCGGGACTTCTCTGTCGACCAAGCAGCAGCAGCAGTACTTGGCAGCCCGCGCG 4877  
Qy 805 SerPro-----SerSerSerProSerProSerProSerPro 816  
Db 4878 TGCCCGCCCTGAAGGACGACTGTTCTGCTCTGCTGCTGCTCTCTTTCGCGCAGCTCG 4937  
Qy 817 SerSerSerProSerSerProSerProSerPro----- 828  
Db 4938 AGCACCGCTGAGTTCGAGCGTGTGTTATCCAGACCCCGGCTTGAGAGAGCGGCTCGAGC 4997  
Qy 829 -----SerProSerProSer----- 833  
Db 4998 TCGTGTGCGACTCCAGCTCGCAACATAGTGTGTCGCGCCGCAACGAGCGGTGACTTCGTG 5057  
Qy 834 ---ProSerArgSerProSerProSerAlaSerProSerProSerSerProSerPro 852  
Db 5058 AGACACGCTCGCGCGCGCTGTCACCGGGTCCAGGGTCCAGGGTCCGCGCTGCCACTGAAGCAGC 5117  
Qy 852 oSerSerSerProSerSerProSerProSerProSerProSer 865  
Db 5118 TCCTGTGTCGAGCGCGCGCGGAGGTGTCGCCCGCAGTCCCGCAGC 5157

## RESULT 14

US-09-917-384-2  
; Sequence 2, Application US/09917384  
; Publication No. US20030096342A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: AUNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOSE  
; FILE REFERENCE: 40170.60501  
; CURRENT APPLICATION NUMBER: US/09/917,384  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 3687  
; TYPE: DNA  
; ORGANISM: Artificial Sequence

; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-384-2

## Alignment Scores:

Pred. No.: 6.37e-50 Length: 3687  
Score: 823.00 Matches: 251  
Percent Similarity: 47.07% Conservatives: 46  
Best Local Similarity: 39.78% Mismatches: 173  
Query Match: 16.03% Indels: 164  
DB: 24 Gaps:

US-09-917-376-1 (1-957) x US-09-917-384-2 (1-3687)

Qy 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445  
Db 307 TGGATGAGCCGATCGTGCATCAACGGCGTCAACGGC-----GGACCC 351  
Qy 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrIleTyrAspSerGlyGlnIleHis 465  
Db 352 GGCCTTGACACATATCTGGACGCGCCCTCTCCAG-----CAGCAGGA 396  
Qy 466 IleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProPro 485  
Db 397 ACCACCCCTGAAGTC-----ATTGAGATTGTCATCTACGATCTG-----CGGGA 441  
Qy 486 SerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspVal 505  
Db 442 CGCAGCTCGCGCGCGCTCGCTCAACGGCGAATGCGCGCT---ACGGCAGCAGGTTG 498  
Qy 506 ThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSerValAsp 525  
Db 499 CAGACCTAT---GAAACGACGATCATCGATTCGATTCGAGTATCTCAGCAATCCGAAG 555  
Qy 526 TyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGln 545  
Db 556 TACTCCAGCTCGCGATCGTGACGATCAT-----GAGCCGAGCTCGCTG 600  
Qy 546 ProAsnAspArgHisAlaValAlaPheSerThrAspGlyGlyLysAsnTyrPheGlnGlySer 565  
Db 601 CCAACGC-GGTAC-----CAATAT 620  
Qy 566 GluProGlyGlyValThrThrGlyThrValAla-----AlaSerAlaAspGly 582  
Db 621 GAGCATTCACGCGTGCACGCGGTCGCGTATT-ACGAGCAAGGTCATCGAGTACGCGC 679  
Qy 583 SerArg-----PheValTyrAlaProGlyAspPro----- 592  
Db 680 TCAGAAATTGACGCCCATTCGGAACGTGTATCATCTACATGAGCGCCGCCACTCGGCT 739  
Qy 593 -----GlyGlnProValVal-----TyrAlaValGlyPheGlyAsnSerTyr 606  
Db 740 GCGTGGGTGGCCCAATAATGCCAGCGGATACGTAAGAGTCCAGAGGTCTCTCAACG 799  
Qy 607 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsn----- 624  
Db 800 CGAGCATCGGGGTCAACGGCATCGAGCTTCGTCCACCAACACGCGGGAATTACACGCCGT 859  
Qy 625 -----ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr 637  
Db 860 TGAAGGAGCCGCTTCATGACCGCCACCCAGAGGTGCGCGGACAGCGGTGGAGTCCGCGA 919  
Qy 638 ArgSerThrAspGlyGlyValThrPheGlnPro----- 648  
Db 920 ATTTCTACCGAGTGAATCTCGACATC-GACGAAGCCGACTACGCGGTGACTTGTACTCG 978  
Qy 649 -----ValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 666  
Db 979 CGGCTCGTCCGCGCTGGCTTCCAGCAGC-----ATCGGATGCTCATCGACACCTTA 1032  
Qy 667 ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 686

Db 1033 -----CGCAAC 1038  
Qy 687 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal----- 703  
Db 1039 GGTGGGGTGGTCCGAACGACCAACAGGCGCCGAGCCGCGGATGTCACACCTTC 1098  
Qy 704 -----Gly 704  
Db 1099 GTCAACAGTCGAAGATTGACCTTCGCGCAGCAGCGCGCTGTGGTGCACCAACGACGGT 1158  
Qy 705 PheGlyValSerAlaProGlySerSerTrpProAlaValPhe----- 718  
Db 1159 GCGGGCTCGCGCAGCGCGCGAGCAAGCCCGACGACTTCGCAACGCGACCTCGAC 1218  
Qy 719 -----ValValGlyThrIleGlyGlyValThrGlyAla----- 729  
Db 1219 GCGTATGTCGATCAAGCGCGCGGTGAGTCGCGACGCGCAAGCGGTGCGAGGATCCG 1278  
Qy 730 -----TyrArgSerAspAsp-----CysGlyThrThrTrpValLeuIleAsnAspAsp 745  
Db 1279 ACAACTGGCAAGAGTCGACCCCATGTGCGACCCGAGGTACACGCG----- 1326  
Qy 746 GlnHisGlnTrpGlyAsnTrpGlyGlnAla-----IleThrGlyAspHis 760  
Db 1327 -----TCGTACGGGTACTGACCAACGCGGTACCGAACTCCCGCATCGCGCGCAGTGG 1380  
Qy 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780  
Db 1381 TTCCCGGCGAGTTTGACCAAGCTTGTGCGAAACGACGCGCGGTGCGAGCTCGAC 1440  
Qy 781 GlyGly-----AlaProSerGlySerProSerProSerValSer 793  
Db 1441 AGTCGAGCGCGCGCTCCGCGCGAGTCCGTGCGCTTCGCGAGTCCGAGCCCGAGT 1500  
Qy 794 ProSerAlaSerProSerLeuSerProSerProSerProSerProSerProSer 813  
Db 1501 CCGAGCCGAGCAGCTCGCCATCGCGCTCGCGCTTCGAGCTCGAGCCCGTCTCGGTG 1560  
Qy 814 Pro-----SerProSerSerProSerProSerProSerProSer 825  
Db 1561 CCGAGCCGAGTCCGAGCCGAGTCCGCGCGAGTCCGCGTCCGAGCTCGAGCCCGTCT 1620  
Qy 826 ProSerProSerProSerProSerProSerProSerProSerProSerProSer 845  
Db 1621 CCGTCCCGAGCCGAGTCCGAGCCGAGTCCGCGTCCGCGTCCGAGCTCGAGC 1680  
Qy 846 ProSerSerProSerProSerProSerProSerProSerProSerProSer 865  
Db 1681 CCGTCTCCGTCGCGAGCCGAGTCCGAGCCGAGTCCGCGTCCGCGTCCGAGC 1740  
Qy 866 SerSerProValSerGlyGlyValValValGlnTyrIleAsnAsnAspSerAlaProGly 885  
Db 1741 TCGTCCCGGTGTCGGGTGCGGTGAGAGGTGAGTACCAAGAACATGATTCGCGCGCGGT 1800  
Qy 886 AsnAsnGlnIleValProGlyLeuGlnValValAsnThrGlySerSerSerValAspLeu 905  
Db 1801 GATAACCAATCAACCGGCTCTCCAGTGTGGTAAATACCGGGTCTGCTCGGTGGATTG 1860  
Qy 906 SerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyr 925  
Db 1861 TCGAGCGTACCGGTGCGGTACTGTTTCCCGGGATGGTGGTGGTGGTGGTGGTGGT 1920  
Qy 926 AsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 945  
Db 1921 AACTGTGACTGGCGCGGATGGGTGGGAATATCCGCGCTCGTTCGGCTCGGTGAAC 1980  
Qy 946 ProAlaThrProThrAlaAspThrTyrLeuGln 956  
Db 1981 CCGGCGACCGCGAGCGGACACCTACCTGACAG 2013

RESULT 15

US-09-917-383-2

; Sequence 2, Application US/09917383

; Publication No. US20030104522A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOXYLICUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,383  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 3687  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-383-2

## Alignment Scores:

Pred. No.:	6.37e-50	Length:	3687
Score:	823.00	Matches:	251
Percent Similarity:	47.07%	Conservative:	46
Best Local Similarity:	39.78%	Mismatches:	173
Query Match:	16.03%	Indels:	164
DB:	10	Gaps:	24

US-09-917-376-1 (1-957) x US-09-917-383-2 (1-3687)

Qy 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445  
Db 307 TGGATGAGCGCATCGCTGCGATCAACGCGCTCAACGCG-----GGACCC 351  
Qy 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrIleTyrAspSerGlyGlyGlnIleHis 465  
Db 352 GCGTTGACGACATATCTGGACGCGCGCTCTCCGAG-----CAGCAGGGA 396  
Qy 466 IleAlaProMetValValGlyLeuGluThrAlaValAsnAspLeuIleSerProPro 485  
Db 397 ACCACCTTGAAGTC-----ATTGAGATTGTCATCTACGATCTG-----CGGGA 441  
Qy 486 SerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspVal 505  
Db 442 CGCGACTGCGCGCGCTCGCTCCCAACGCGAACTGCGCGCT---ACGCGACGAGGTTG 498  
Qy 506 ThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAsp 525  
Db 499 CAGACCTAT---GAAACGACGATCATCGATTCGATTCGAGGATTCCTGAGCAATCCGAAG 555  
Qy 526 TyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGln 545  
Db 556 TACTCCAGCTCGGATCGTACCATCT---GAGCGGACTCGCTG 600  
Qy 546 ProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySer 565  
Db 601 CCAACGCG-GGTAC-----CAATAT 620  
Qy 566 GluProGlyValThrThrGlyThrValAla-----AlaSerAlaAspGly 582  
Db 621 GAGCATTCAGCGGTGTCACGCGCGTCCGCTATT-ACGAGCAAGGCTACGAGTACGCGC 679  
Qy 583 SerArg-----PheValTrpAlaProGlyAspPro--- 592  
Db 680 TCAGAAATTCACGCCCATTCGAGCGTGTACATCTATATGAGCGCGCCACTCGGGCT 739  
Qy 593 -----GlyGlnProValVal-----TyrAlaValGlyPheGlyAsnSerTrp 606  
Db 740 GCGTTGGTGGCCCAATAATGCCAGCGGATACGTACAGGAGTCCAGAAGGCTCTCAACG 799



```
QY 607 AlaAlaSerGlnGlyValProAlaAlaAsnAlaGlnIleArgSerAspArgValAsn----- 624
Db 800 CGAGCATCGGGTCAACGGCATCGAGCGCTTCGTACCAACACGGCGAATTACACGCGGT 859
QY 625 -----ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr 637
Db 860 TGAAGAGCGGTCATGACCGCCACCCAGCAGGTCCGCGGACAGCCGGTGGAGTCGCGCA 919
QY 638 ArgSerThrAspGlyGlyValThrPheGlnPro----- 648
Db 920 ATTCTTACCAGTGGATCCTGACATC-GACGAAGCGGACTACGCGGTGGACTTGTACTCG 978
QY 649 -----ValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 666
Db 979 CGGCTCGTCGCCCGCTGCTTTTCAAGCAGC-----ATCGGCATGTCTCATCAGACACCTTA 1032
QY 667 ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 686
Db 1033 -----CGCAAC 1038
QY 687 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal----- 703
Db 1039 GGTGGGGTGGTCCGAACGACCAACAGGCCCGGACACCGCGCCGATGTCAACACCTTC 1098
QY 704 -----Gly 704
Db 1099 GTCAACAGTCGAAGATTGACCTTCGGCAGCAGCGCGCTGTGTGCAACAGAACGGT 1158
QY 705 PheGlyLysSerAlaProGlySerSerTyrProAlaValPhe----- 718
Db 1159 GCGGGCTCGCGCAGCGCGCAGCAAGCCCGGACCGGACTTCCCGAACGCGCACCTCGAC 1218
QY 719 -----ValValGlyThrIleGlyGlyValThrGlyAla----- 729
Db 1219 GCGTATGTCGGATCAAGCGCGCGGTGAGTCGGACGCGCAAGCGCTGGAGCGATCCG 1278
QY 730 -----TyrArgSerAspAsp-----CysGlyThrThrTrpValLeuIleAsnAspAsp 745
Db 1279 ACAACTGGCAAGAGTCGGACCCCATGTGCGACCGCGAGCTACACGAGC----- 1326
QY 746 GlnHisGlnTyrGlyAsnTrpGlyGlnAla-----IleThrGlyAspHis 760
Db 1327 -----TCGTACGGGGTACTGACCAACGCGGTTACCGAACTCCCGCATCGCGCGCAGTGG 1380
QY 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
Db 1381 TTCCCGCGCGAGTTTGACAGCTGTTCGCGAAGCAGCAGCGCGGTCGCGAGCTCGAC 1440
QY 781 GlyGly-----AlaProSerGlySerProSerProSerValSer 793
Db 1441 AGCTCGAGCCCGCGCTCCGCGCGAGTCGCTCGGCTTCGCGAGTCCGAGCCCGAGT 1500
QY 794 ProSerAlaSerProSerLeuSerProSerProSerProSerSerSerProSerProSer 813
Db 1501 CCGAGCCCGAGCAGCTCGCATCGCGTCCGCGTCTCCGAGCTCGAGCCCGTCTCCGTCG 1560
QY 814 Pro-----SerProSerSerSerProSerSerProSer 825
Db 1561 CCGAGCCCGAGTCCGAGCCCGAGTAGCTCGCCGTCGCGCTCCGAGCTCGAGCCCGTCT 1620
QY 826 ProSerProSerProSerProSerArgSerProSerProSerAlaSerProSer 845
Db 1621 CCGTCGCGGAGCCCGAGTCCGAGCCCGAGTAGCTCGCGCTCGCGTCTCCGAGCTCGAGC 1680
QY 846 ProSerSerSerProSerProSerSerSerProSerSerProSerProSerProSer 865
Db 1681 CCGTCTCCGTCGCGAGCCCGAGTCCGAGCCCGAGTAGCTCGCGTCTCCGAGCTCGAGC 1740
QY 866 SerSerProValSerGlyGlyValLysValGlnTyrIlyAsnAsnAspSerAlaProGly 885
Db 1741 TCGTCGCGGTGTGCGGTGGGCTGAAGGTGAGTACAAAGAACATGATTCGCGCGCGGT 1800
QY 886 AspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySerSerValAspLeu 905
```

Search completed: May 15, 2004, 03:52:53  
Job time : 13740 sec